
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jul 7 15:32:14 2000; MasPar time 3.22 Seconds
Tabular output not generated. 110.221 Million cell updates/sec

Title: >US-09-300-612-1
Description: (1-15) from US09300612.pep
Perfect Score: 116
Sequence: 1 LKAMDPTPLWIKTE 15

Scoring table: PAM 150
Gap 15

Searched: 189963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 19.457; Variance 66.525; scale 0.292

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	116	100.0	15	1	W11575	8.50e-05
2	116	100.0	15	1	N-terminal peptide from	8.50e-05
3	73	62.9	10	1	W53841	8.50e-05
4	72	62.1	1189	1	N-terminus of opossum	3.07e-00
5	72	62.1	1213	1	TARA-binding protein-a	3.87e-00
6	72	62.1	1213	1	Drosophila TARA-binding	3.87e-00
7	60	51.7	222	1	TATA-binding protein a	3.87e-00
8	60	51.7	222	1	Human 5' EST secreted	5.78e-01
9	60	51.7	222	1	Human 5' EST secreted	5.78e-01
10	60	51.7	222	1	Human secreted protein	5.78e-01
11	60	51.7	222	1	Homo sapiens clone AM4	5.78e-01
12	60	51.7	222	1	Human 5' EST secreted	5.78e-01
13	56	48.3	97	1	Secreted protein encod	5.78e-01
14	55	47.4	311	1	Tobacco leaf curling v	1.38e-02
15	55	47.4	445	1	Human peroxisomal thio	1.71e-02
16	55	47.4	551	1	H. pylori cytoplasmic	1.71e-02
17	55	47.4	594	1	Human cystathionine be	1.71e-02
18	55	47.4	769	1	H. pylori cytoplasmic	1.71e-02
19	54	46.6	233	1	Env gene of simian imm	2.11e-02
20	54	46.6	395	1	Human serine protease	2.11e-02
21	54	46.6	544	1	Tetracycline resistanc	2.11e-02
22	54	46.6	647	1	Protein encoded by ORF	2.11e-02
23	54	46.6	986	1	Human HPD8V78 protein	2.11e-02
					Protein tyrosine-kinas	2.11e-02

24	54	46.6	986	1	R85091	2.11e+02
25	54	46.6	1891	1	W52844	2.11e+02
26	54	46.6	3413	1	A. mediterranei rifam	2.11e+02
27	53	45.7	222	1	Human 5' EST secreted	2.61e+02
28	53	45.7	1294	1	W5798	2.61e+02
29	53	45.7	1321	1	R88423	2.61e+02
30	53	45.7	1622	1	W60163	2.61e+02
31	53	45.7	4545	1	W22611	2.61e+02
32	53	45.7	4550	1	W22606	2.61e+02
33	53	45.7	4550	1	W23716	2.61e+02
34	52	44.8	268	1	W98880	3.22e+02
35	52	44.8	1434	1	R94380	3.22e+02
36	52	44.8	1434	1	W52199	3.22e+02
37	52	44.8	1434	1	W72968	3.22e+02
38	52	44.8	1447	1	R75375	3.22e+02
39	52	44.8	1447	1	W52205	3.22e+02
40	52	44.8	1618	1	R27200	3.22e+02
41	52	44.8	1621	1	R60127	3.22e+02
42	52	44.8	1621	1	W60162	3.22e+02
43	52	44.8	1688	1	W52848	3.22e+02
44	52	44.8	2986	1	R44432	3.22e+02
45	52	44.8	3164	1	R94345	3.22e+02

ALIGNMENTS

RESULT 1
ID W11575 standard; peptide; 15 AA.
AC W11575;
DE 20-MAR-1997 (first entry)
KW Lethal toxin neutralising factor; LTNF; opossum; bee toxin;
KW scorpion toxin; plant toxin; bacterial toxin; venom; sting;
KW snake bite.
OS Didelphis virginiana.
PN US5576297-A.
PD 19-NOV-1996.
PF 10-MAY-1993; 058387.
PR 10-MAY-1993; US-058387.
PR 22-SEP-1994; US-310340.
PA (LIPP/) LIPPS B V.
PA (LIPP/) LIPPS F W.
PI Lipps BV, Lipps FW;
WPI: 97-011287/01.
PT Treatment of victims of bee or scorpion stings or plant or bacterial
toxins - by admin. of lethal toxin-neutralising factor or its
N-terminal peptide
PS Claim 7; Column 9; 9pp; English.
CC The present sequence is from the N-terminus of a 68 kD protein
CC purified from the serum of the opossum *Didelphis virginiana*. The
CC full-length protein is a lethal toxin neutralising factor (LTNF).
CC The use of purified LTNF or of the chemically synthesised 15mer
CC N-terminal peptide for treating victims of bee stings, scorpion
CC stings and bacterial or plant toxins is claimed. The patent
CC disclosure does not provide any evidence for neutralising activity
CC against these various toxins. There is evidence of significant
CC neutralising activity of the opossum LTNF and the 15mer peptide
CC against venom from snakes of the families Crotalidae, Elapidae,
CC Hydroliidae and Viperidae.
SQ Sequence 15 AA;

Query Match 100.0%; Score 116; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 8.50e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LKAMDPTPLWIKTE 15
QY 1 LKAMDPTPLWIKTE 15

RESULT 2
ID W53841 standard; peptide; 15 AA.
AC W53841;

DT 08-JUL-1998 (first entry)
 DE N-terminus of opossum LTNF.
 KW LTNF; lethal toxin neutralising factor; opossum; envenomation; therapy;
 KW anti-haemorrhagic protein; Elapidae; Viperidae; sea snake; snake bite;
 KW sepsis; allergy; bee sting; scorpion sting; plant toxin; bacterial toxin;
 KW histamine reaction treatment.
 OS Didelphis virginiana.
 PN US5744449-A.
 PD 28-APR-1998.
 PF 03-JUN-1996; 657163.
 PR 03-JUN-1996; US-657163.
 PR 10-MAY-1993; US-058387.
 PR 22-SEP-1994; US-310340.
 PA (LIPP/) LIPPS B V.
 PA (LIPP/) LIPPS F W.
 PI LIPPS BV, LIPPS FW;
 PI WPI; 98-271108/24.
 PT Lethal Toxin Neutralising Factor peptide from opossum - can
 neutralise venom(s) from all major families of poisonous snakes
 Claim 1: Column 11; lipp; English.
 CC This sequence represents the peptide of the invention. It is a Lethal
 CC Toxin Neutralising Factor (LTNF) moiety from a 68 kDa anti-haemorrhagic
 CC protein derived from an opossum. The peptide can be used in a method for
 CC treating a victim of envenomation from a poisonous snake, preferably a
 CC poisonous snake from the family of Elapidae, Viperidae or sea snake.
 CC It is useful for the treatment of snake bites, sepsis, allergies caused
 CC by the environment and treatment of bee or scorpion stings or toxicities
 CC caused by plant or bacterial toxins. The peptide can also be used in
 CC histamine reaction treatment. The peptide can be used in envenomation
 CC treatment for a variety of snakes without prior identification of the
 CC snake. Being short it can be synthetically prepared rather than the
 CC current production in horses, where some people can show hypersensitivity
 CC to horse proteins.
 SQ Sequence 15 AA;
 Query Match 100.0%; Score 116; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 8.50e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 LKAMDTPPLWIKTE 15
 QY |||||
 RESULT 3
 ID W53843 standard; peptide; 10 AA.
 AC W53843;
 DE 08-JUL-1998 (first entry)
 KW LTNF; lethal toxin neutralising factor; opossum; envenomation; therapy;
 KW anti-haemorrhagic protein; Elapidae; Viperidae; sea snake; snake bite;
 KW sepsis; allergy; bee sting; scorpion sting; plant toxin; bacterial toxin;
 KW histamine reaction treatment.
 OS Didelphis virginiana.
 PN US5744449-A.
 PD 28-APR-1998.
 PF 03-JUN-1996; 657163.
 PR 03-JUN-1996; US-657163.
 PR 10-MAY-1993; US-058387.
 PR 22-SEP-1994; US-310340.
 PA (LIPP/) LIPPS B V.
 PA (LIPP/) LIPPS F W.
 PI LIPPS BV, LIPPS FW;
 PI WPI; 98-271108/24.
 PT Lethal Toxin Neutralising Factor peptide from opossum - can
 neutralise venom(s) from all major families of poisonous snakes
 Claim 1: Column 11; lipp; English.
 CC This sequence represents the peptide of the invention. It is a Lethal
 CC Toxin Neutralising Factor (LTNF) moiety from a 68 kDa anti-haemorrhagic
 CC protein derived from an opossum. The peptide can be used in a method for
 CC treating a victim of envenomation from a poisonous snake, preferably a
 CC poisonous snake from the family of Elapidae, Viperidae or sea snake.
 CC It is useful for the treatment of snake bites, sepsis, allergies caused

CC by the environment and treatment of bee or scorpion stings or toxicities
 CC caused by plant or bacterial toxins. The peptide can also be used in
 CC histamine reaction treatment. The peptide can be used in envenomation
 CC treatment for a variety of snakes without prior identification of the
 CC snake. Being short it can be synthetically prepared rather than the
 CC current production in horses, where some people can show hypersensitivity
 CC to horse proteins.
 SQ Sequence 10 AA;
 Query Match 62.9%; Score 73; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 3.07e-00;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 LKAMDTPPL 10
 QY |||||
 RESULT 4
 ID R56496 standard; Protein; 1189 AA.
 AC R56496;
 DE 23-MAR-1995 (first entry)
 KW TATA-binding protein-associated factor dTAFI50.
 KW TATA-binding protein associated factor; dTAFI50; screening;
 KW diagnostic; therapeutic; gene transcription regulation.
 OS Drosophila.
 FH Key Location/Qualifiers
 FT misc_difference 923
 FT /note= "Val or Leu"
 FT misc_difference 1106
 FT /note= "Arg, Pro or His"
 FT misc_difference 1172
 FT /note= "STOP"
 FT misc_difference 1176
 FT /note= "STOP"
 PN W09417087-A.
 PD 04-AUG-1994.
 PF 28-JAN-1994; U01114.
 PR 28-JAN-1993; US-013412.
 PR 30-JUN-1993; US-087119.
 PA (REGC) UNIV CALIFORNIA.
 PI Conai L, Dyalact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 PI WPI; 94-264019/32.
 DR P-PSDB; q70733.
 PT TATA-binding protein associated protein factors - and
 PT corresponding nucleotide sequence and deriv. antibodies, useful
 PT in screening, diagnostics and therapeutics
 PS Disclosure; Page 156; 180pp; English.
 CC The TATA-binding protein associated factor hTAFI50 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.
 SQ Sequence 1189 AA;
 Query Match 62.1%; Score 72; DB 1; Length 1189;
 Best Local Similarity 53.3%; Pred. No. 3.87e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Db 589 LSAMDDSPVLWIRLD 603
 QY |||||
 RESULT 5
 ID W06086 standard; Protein; 1213 AA.
 AC W06086;
 DE 27-JAN-1997 (first entry)
 KW Drosophila TATA-binding protein associated factor dTAFI150 protein.
 KW Drosophila; TATA-binding protein; TBP associated factor; TFIID;
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;

KW holoenzyme; lambda-gt11; expression library.
 OS Drosophila melanogaster.
 PN US5534410-A.
 PD 09-JUL-1996.
 PF 28-JAN-1993; 013412.
 PR 28-JAN-1993; US-013412.
 PR 30-JUN-1993; US-087119.
 PR 28-JAN-1994; US-188582.
 PA (REGC) UNIV CALIFORNIA.
 PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 DR WPI: 96-333245/33.
 DR N-PSDB; T42219.
 DR Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 PS Examples: Column 123-132; 86pp; English.
 CC This is the amino acid sequence of the Drosophila TATA-binding protein
 CC (TBP) associated factor (TAF) designated TAFII160. The protein is a
 CC component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 60 kD by SDS-PAGE.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.
 CC Sequence 1213 AA;
 SQ

Query Match 62.1%; Score 72; DB 1; Length 1213;
 Best Local Similarity 53.3%; Pred. No. 3.87e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 620 LSAMDSPVLWIRLD 634
 QY 1 LKAMDTPPLWKTE 15

RESULT 6
 ID W25029 standard; Protein; 1213 AA.
 AC W25029;
 DT 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, dTAFII150.
 KW TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW Initiation.
 OS Drosophila sp.
 PN US5637686-A.
 PD 10-JUN-1997.
 PF 28-JAN-1993; 013412.
 PR 28-JAN-1993; US-013412.
 PR 30-JUN-1993; US-087119.
 PR 28-JAN-1994; US-188582.
 PR 09-MAY-1996; US-646715.
 PA (REGC) UNIV CALIFORNIA.
 PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 DR WPI: 97-319113/29.
 DR N-PSDB; T79605.
 DR Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 PS Example 1; Column 131-138; 86pp; English.
 CC W25029 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, dTAFII150 (mol. weight 150KD). TAF peptides derived
 CC from dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80,
 CC dTAFII110, dTAFII150, and dTAFII250, their human equivalents and
 CC nucleic acids encoding them, are used to modulate transcription,

CC including transcription initiation. TAFs are nuclear proteins involved
 CC in RNA polymerase I, II and III transcription. The peptides act by
 CC binding to a different TAF, an activator or TBP (TATA-binding protein)
 CC or competitively inhibiting association of a TAF domain with another
 CC compound, typically a protein like TBP or another TAF, an activator,
 CC or DNA.
 SQ Sequence 1213 AA;

Query Match 62.1%; Score 72; DB 1; Length 1213;
 Best Local Similarity 53.3%; Pred. No. 3.87e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 620 LSAMDSPVLWIRLD 634
 QY 1 LKAMDTPPLWKTE 15

RESULT 7
 ID Y12680 standard; peptide; 222 AA.
 AC Y12680;
 DT 21-JUN-1999 (first entry)
 DE Human 5' EST secreted protein.
 KW human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductuve hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 OS Homo sapiens.
 PN W09906549-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998; IB1231.
 PR 01-AUG-1997; US-905279.
 PA (GEST) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 PI WPI: 99-153779/13.
 DR N-PSDB; X51449.
 DR New nucleic acids encoding human secreted proteins - obtained from
 PT CDNA libraries derived from testis, ovary, uterus and spleen tissue
 PS Example 28; Page 159-160; 522pp; English.
 CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y12681 to Y12913,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haematopoiesis regulating activity, tissue growth regulating
 CC activity, reproductuve hormone regulating activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity
 CC or other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell. This sequence represents an
 CC oligonucleotide used in an example in the invention, to the isolate the
 CC 5' EST sequences of the invention.
 SQ Sequence 222 AA;

Query Match 51.7%; Score 60; DB 1; Length 222;
 Best Local Similarity 60.0%; Pred. No. 5.78e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 135 PMDPSVPIWI 144
 QY 3 AMDPTPLWI 12

RESULT 8
 ID Y04156 standard; Protein; 222 AA.
 AC Y04156;

DT 16-JUN-1999 (first entry)
 DE Human 5' EST secreted protein SQ ID NO:27.
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductively hormone regulation; chemotactic; chemokine; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 OS Homo sapiens.
 PN W0906439-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998; IB1233.
 PR 01-AUG-1997; US-904468.
 PA (GENSET) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 DR N-PSDB; X19983.
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT CDNA libraries derived from liver, lung, large intestine, colon,
 PT thyroid and pancreas tissue
 PS Example 28; Page 157-158; 398pp; English.
 CC X40251 to X40397 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y11533 to Y11679,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haematopoiesis regulating activity, tissue growth regulating
 CC activity, reproductively hormone regulating activity, chemotactic/
 CC chemokine activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell.
 CC This sequence represents the human secreted protein encoded by 5' EST
 CC Protein from a 5' EST from an example of the present invention.
 SQ Sequence 222 AA;

Query Match 51.7%; Score 60; DB 1; Length 222;
 Best Local Similarity 60.0%; Pred. No. 5.78e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 135 PMDPSVPIWI 144
 QY :|||: |||
 3 AMDPTPLWI 12

RESULT 9
 ID Y12986 standard; Protein; 222 AA.
 AC Y12986;
 DT 22-JUN-1999 (first entry)
 DE Human secreted protein encoded by 5' EST clone 58-35-2-F10-FL2.
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductively hormone regulation; chemotactic; chemokine; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 OS Homo sapiens.
 PN W0906552-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998; IB1236.
 PR 01-AUG-1997; US-905223.
 PA (GENSET) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 DR N-PSDB; X51777.
 PT New isolated brain-derived nucleic acids - used to develop products
 PT which may have cytokine, immune, regulatory, haematopoiesis
 PT regulating, anti-inflammatory or tumour inhibition activity

PS Example 28; Page 159-160; 577pp; English.
 CC X51787 to X52019 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y12987 to Y13219,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haematopoiesis regulating activity, tissue growth regulating
 CC activity, reproductively hormone regulating activity, chemotactic/
 CC chemokine activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity
 CC or other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell.
 CC This sequence represents the human secreted protein encoded by 5' EST
 CC clone 58-35-2-F10-FL2.
 SQ Sequence 222 AA;

Query Match 51.7%; Score 60; DB 1; Length 222;
 Best Local Similarity 60.0%; Pred. No. 5.78e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 135 PMDPSVPIWI 144
 QY :|||: |||
 3 AMDPTPLWI 12

RESULT 10
 ID W29670 standard; Protein; 222 AA.
 AC W29670;
 DT 09-NOV-1998 (first entry)
 DE Homo sapiens clone AM42_3 secreted protein.
 KW Clone; secreted protein.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide
 FT 2..14
 FT /note= "signal peptide"
 FN W09832853-A2.
 PD 30-JUL-1998.
 PF 23-JAN-1998; U01396.
 PR 24-JAN-1997; US-788789.
 PA (GEMY) GENETICS INST INC.
 PI Agostino MT, Jacobs K, Lavallie ER, Mccoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 98-427949/36.
 DR N-PSDB; V40540.
 PT New isolated polynucleotide(s) and secreted proteins - isolated from
 PT human foetal kidney, adult brain, adult salivary gland, foetal brain
 PT and adult testes cDNA libraries
 PS Claim 15; Page 65-66; 109pp; English.
 CC The sequence is that of a secreted protein. Such a
 CC protein can have biological activities, e.g. nutritional
 CC activity, cytokine and cell proliferation/differentiation activity,
 CC immune stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokine activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, tumour inhibition activity, and other
 CC activities.
 SQ Sequence 222 AA;

Query Match 51.7%; Score 60; DB 1; Length 222;
 Best Local Similarity 60.0%; Pred. No. 5.78e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 135 PMDPSVPIWI 144
 QY :|||: |||
 3 AMDPTPLWI 12

```

RESULT 11
ID W93620 standard; Protein; 222 AA.
AC W93620;
DE 21-JUN-1999 (first entry)
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductively hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9906551-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1235.
PR 01-AUG-1997; US-905133.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153781/13.
PT New nucleic acids encoding human secreted - proteins obtained from
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals
PT and fetal brain tissue
PS Example 28; Page 157-158; 434pp; English.
CC X39440 to X39597 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y11374 to Y11531,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductively hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell. This represents a human 5' EST
CC secreted protein encoded by X39430.
SQ Sequence 222 AA;

Query Match 51.7%; Score 60; DB 1; Length 222;
Best Local Similarity 60.0%; Pred. No. 5.78e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 135 PMDPSVPIWI 144
QY :|||: |||
3 AMDPTPLWI 12

RESULT 12
ID Y01594 standard; Protein; 222 AA.
AC Y01594;
DE 18-JUN-1999 (first entry)
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductively hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9906554-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1238.
PR 01-AUG-1997; US-905134.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153784/13.

```

```

DR N-PSDB; X26672.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
PS Example 28; Page 161-161; 622pp; English.
CC The present sequence is encoded by an extended cDNA sequence derived
CC from a 5' EST encoding a secreted protein. The specification describes
CC 5' expressed sequence tags (ESTs, see X40826-X41093) for human secreted
CC proteins (see Y01602 and Y1194-Y12260). The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein.
CC The nucleic acid sequences can be used for producing secreted human
CC gene products. They can also be used to develop products for diagnosis
CC and therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductively hormone
CC and thrombolytic activity, chemotactic/chemokinetic activity, haemostatic
CC and tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into a
CC cell.
SQ Sequence 222 AA;

Query Match 51.7%; Score 60; DB 1; Length 222;
Best Local Similarity 60.0%; Pred. No. 5.78e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 135 PMDPSVPIWI 144
QY :|||: |||
3 AMDPTPLWI 12

RESULT 13
ID W56498 standard; Protein; 97 AA.
AC W56498;
DE 03-SEP-1998 (revised)
DT 11-AUG-1998 (first entry)
DE Tobacco leaf curling virus gene protein C4.
KW Tobacco leaf curling virus gene; TLCV; promoter; C4 protein.
OS Tobacco leaf curling virus.
PN J10070982-A.
PD 17-MAR-1998.
PF 30-AUG-1996; 230394.
PR 30-AUG-1996; JP-230394.
PA (NORQ ) NORINSUISANSRO KYUSHU NOGTO SHIKENJOYO.
DR WPI; 98-233630/21.
PT Tobacco leaf curling virus gene - useful for inserting into vectors
PT for expression in, e.g. tomato plants
PS Example 4; Fig 5; 9pp; Japanese.
CC This sequence represents the C4 protein encoded by the tobacco leaf
CC curling virus (TLCV) gene of the invention. TLCV gene or its promoter can
CC be inserted into a vector for expression in plants, e.g. tobacco and
CC tomato. This sequence is believed to be encoded by the TLCV gene shown in
CC V29761.
SQ Sequence 97 AA;

Query Match 48.3%; Score 56; DB 1; Length 97;
Best Local Similarity 53.3%; Pred. No. 1.38e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 42 LKAROMSSPTWKTE 56
QY :|||: :|||
1 LKAMDPTPLWKTE 15

RESULT 14
ID W58530 standard; Protein; 311 AA.
AC W58530;
DR 01-SEP-1998 (first entry)

```

RESULT	15	
ID	W20095	standard; Protein; 446 AA.
AC	W20095;	
DE	08-JUL-1997	(first entry)
DT	H. pylori cytoplasmic protein 10677187.aa.	
DE	Cytoplasmic; vaccine; prevention; treatment; infection; identification;	
KW	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;	
KW	duodenal ulcer disease; chronic gastritis; diagnosis; envelope.	
KW	Helicobacter pylori.	
OS	Helicobacter pylori.	
PN	W09640893-Al.	
PD	19-DEC-1996.	
PF	06-JUN-1996; U09122.	
PF	07-JUN-1995; US-487032.	
PR	01-APR-1996; US-630405.	
PA	(ASTR) ASTRA AB.	
PI	Berglindh OT, Smith D, Mellgaard BL;	
DR	WPI; 97-052306/05.	
DR	N-PSDB; T67345.	
PPT	Helicobacter pylori nucleic acid sequences and related	
PPT	polypeptide(s) - useful for vaccines to treat or prevent H. pylori	
PPT	infection, and to detect Helicobacter	
PPT	Claim 61; pages 329-330; 1481pp; English.	
CC	This sequence is a H. pylori cytoplasmic protein involved in	
CC	genome replication, transcription, recombination and repair.	
CC	The protein may be used in a vaccine to prevent or treat H. pylori	
CC	infection or to identify H. pylori polypeptide binding compounds.	

(TM)

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.	
		Match							
1	116	100.0	15	1	US-08-657-	Sequence 1,	Applicatio	1.05e-04	
2	116	100.0	15	1	US-08-310-	Sequence 1,	Applicatio	1.05e-04	
3	73	62.9	10	1	US-08-657-	Sequence 2,	Applicatio	2.36e+00	
4	72	62.1	1213	1	US-08-646-	Sequence 20,	Applicati	2.95e+00	
5	72	62.1	1213	1	US-08-188-	Sequence 20,	Applicati	2.95e+00	
6	57	49.1	206	2	US-08-477-	Sequence 18,	Applicati	7.51e+01	
7	57	49.1	355	2	US-08-846-	Sequence 93,	Applicati	7.51e+01	
8	56	48.3	627	4	PCR-US95-1	Sequence 10,	Applicati	9.25e+01	
9	56	48.3	627	1	US-08-291-	Sequence 10,	Applicati	9.25e+01	
10	55	47.4	311	2	US-09-100-	Sequence 1,	Applicatio	1.14e+02	
11	55	47.4	311	1	US-08-872-	Sequence 1,	Applicatio	1.14e+02	
12	55	47.4	351	1	US-08-120-	Sequence 2,	Applicatio	1.14e+02	
13	54	46.6	196	4	PCR-US93-0	Sequence 7,	Applicatio	1.40e+02	
14	54	46.6	196	1	US-08-063-	Sequence 7,	Applicatio	1.40e+02	
15	54	46.6	222	1	US-08-035-	Sequence 21,	Applicati	1.40e+02	
16	54	46.6	396	2	US-08-814-	Sequence 4,	Applicatio	1.40e+02	
17	54	46.6	396	2	US-08-944-	Sequence 4,	Applicatio	1.40e+02	
18	54	46.6	396	2	US-08-850-	Sequence 4,	Applicatio	1.40e+02	
19	54	46.6	986	4	PCR-US95-0	Sequence 15,	Applicati	1.40e+02	
20	54	46.6	986	2	US-08-702-	Sequence 15,	Applicati	1.40e+02	
21	54	46.6	986	2	US-08-449-	Sequence 15,	Applicati	1.40e+02	
22	54	46.6	1104	4	PCR-US95-0	Sequence 36,	Applicati	1.40e+02	
23	54	46.6	1104	1	US-08-222-	Sequence 36,	Applicati	1.40e+02	

CC TITLE OF INVENTION: EMBODIMENTS OF NATURAL AND
CC TITLE OF INVENTION: SYNTHETIC LNFS AND THEIR
CC TITLE OF INVENTION: UTILITY AS TREATMENT FOR ENVENOMATION
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BINIE V. LIPPS
CC STREET: 4509 MIMOSA DR.
CC CITY: BELLAIRE
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77401
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" FLOPPY DISK, 1.44 MB
CC COMPUTER: IBM COMPATIBLE
CC OPERATING SYSTEM: MS-DOS 5.0/WINDOWS 3.1
CC SOFTWARE: MS WORD 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/657,163A
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/310,340
CC FILING DATE: 22 SEPTEMBER 1994
CC CLASSIFICATION: 514
CC APPLICATION NUMBER: 08/058,387
CC FILING DATE: 10 MAY 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: JOHN R. CASPERSON
CC REGISTRATION NUMBER: 28,198
CC REFERENCE/DOCKET NUMBER: FWL-PAT-US-011
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 713-482-2961
CC TELEFAX: 713-663-7290
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 10
CC TYPE: AMINO ACID
CC STRANDEDNESS: SINGLE
CC TOPOLOGY: LINEAR
CC MOLECULE TYPE: PEPTIDE IN SEQ ID NO: 2
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N
CC ORIGINAL SOURCE: SYNTHETIC
CC SEQUENCE 10 AA: 1082 MW; 684 CN;

Query Match 62.9%; Score 73; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.36e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LKAMDTPPL 10
QY 1 LKAMDTPPL 10

RESULT 4
ID US-08-646-715-20 STANDARD; PRT; 1213 AA.

AC xxxxxx
XX
DT
XX
DE Sequence 20, Application US/08646715
XX
CC Sequence 20, Application US/08646715
CC Patent No. 5637686
CC GENERAL INFORMATION:
CC APPLICANT: Tjian, Robert
CC APPLICANT: Comai, Lucio
CC APPLICANT: Dynlacht, Brian D.
CC APPLICANT: Hoey, Timothy
CC APPLICANT: Ruppert, Siegfried

CC APPLICANT: Tanese, Naoko
CC APPLICANT: Wang, Edith
CC APPLICANT: Weinzierl, Robert O.J.
CC TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
CC TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
CC NUMBER OF SEQUENCES: 36
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/646,715
CC FILING DATE: 09-MAY-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/188,582
CC FILING DATE: 28-JAN-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard A
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC TELEX: 910 277299
CC INFORMATION FOR SEQ ID NO: 20:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1213 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1213 AA: 138533 MW; 7456135 CN;

Query Match 62.1%; Score 72; DB 1; Length 1213;
Best Local Similarity 53.3%; Pred. No. 2.95e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 620 LSAMDDSPVLWRLD 634
QY 1 LKAMDTPPLWKTE 15

RESULT 5
ID US-08-188-582-20 STANDARD; PRT; 1213 AA.

AC xxxxxx
XX
DT
XX
DE Sequence 20, Application US/08188582
XX
CC Sequence 20, Application US/08188582
CC Patent No. 5534410
CC GENERAL INFORMATION:
CC APPLICANT: Tjian, Robert
CC APPLICANT: Comai, Lucio
CC APPLICANT: Dynlacht, Brian D.
CC APPLICANT: Hoey, Timothy
CC APPLICANT: Ruppert, Siegfried
CC APPLICANT: Tanese, Naoko
CC APPLICANT: Wang, Edith
CC APPLICANT: Weinzierl, Robert O.J.
CC TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
CC TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
CC NUMBER OF SEQUENCES: 36

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/188,582
CC FILING DATE: 28-JAN-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard A.
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC TELEX: 910 277299
CC INFORMATION FOR SEQ ID NO: 20:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1213 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1213 AA; 138533 MW; 7456135 CN;
SQ
Query Match 62.1%; Score 72; DB 1; Length 1213;
Best Local Similarity 53.3%; Pred. No. 2.95e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 620 LSAMDSPVLWIRLD 634
QY 1 LKAMDPTPLWKTE 15
RESULT 6
ID US-08-477-396A-18 STANDARD; PRT; 206 AA.
XX XXXXXX
AC XXXXXX
DT
DE
Sequence 18, Application US/08477396A
XX
Sequence 18, Application US/08477396A
CC Patent No. 5872235
CC GENERAL INFORMATION:
CC APPLICANT: Chen, Lan Bo
CC APPLICANT: Bao, Shideng
CC APPLICANT: Liu, Yuan
CC TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
CC TITLE OF INVENTION: ISOLATING SAME
CC NUMBER OF SEQUENCES: 19
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
CC STREET: Ten Post Office Square
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/477,396A

CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/146,488
CC FILING DATE: 29-OCT-1993
CC APPLICATION NUMBER: US 08/448,388
CC FILING DATE: 28-MAY-1996
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/12502
CC FILING DATE: 31-OCT-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Heine, Holliday C.
CC REGISTRATION NUMBER: 34,346
CC REFERENCE/DOCKET NUMBER: DFCI-333BX
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-2290
CC TELEFAX: (617) 451-0313
CC INFORMATION FOR SEQ ID NO: 18:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 206 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
CC SEQUENCE 206 AA; 24075 MW; 214013 CN;
SQ
Query Match 49.1%; Score 57; DB 2; Length 206;
Best Local Similarity 66.7%; Pred. No. 7.51e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 82 LDGNPPLWI 90
QY 4 MDPTPLWI 12
RESULT 7
ID US-08-846-762-93 STANDARD; PRT; 355 AA.
XX XXXXXX
AC XXXXXX
DT
DE
Sequence 93, Application US/08846762A
XX
Sequence 93, Application US/08846762A
CC Patent No. 5994072
CC GENERAL INFORMATION:
CC APPLICANT: Lam, Joseph S.
CC APPLICANT: Burrows, Lori
CC APPLICANT: Charter, Deborah
CC APPLICANT: de Kievit, Teresa
CC TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assem
CC TITLE OF INVENTION: of O-Antigen in Pseudomonas aeruginosa
CC FILE REFERENCE: 6580-089
CC CURRENT APPLICATION NUMBER: US/08/846,762A
CC CURRENT FILING DATE: 1997-04-30
CC NUMBER OF SEQ ID NOS: 100
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 93
CC LENGTH: 355
CC TYPE: PRT
CC ORGANISM: Pseudomonas aeruginosa
CC SEQUENCE 355 AA; 40070 MW; 672413 CN;
SQ
Query Match 49.1%; Score 57; DB 2; Length 355;
Best Local Similarity 50.8%; Pred. No. 7.51e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 236 PMNPVTALWI 245
QY 1 LKAMDPTPLWKTE 15

QY 3 AMDPTPLWI 12

RESULT 8
ID PCT-US95-10579-10 STANDARD: PRT: 627 AA.
XX
XX
XX
DT
XX
Sequence 10, Application PC/TUS9510579
XX
Sequence 10, Application PC/TUS9510579
XX
GENERAL INFORMATION:
CC APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION
CC TITLE OF INVENTION: DNA ENCODING A HUMAN BETAINE/GABA
CC TITLE OF INVENTION: TRANSPORTER AND USES THEREOF
CC NUMBER OF SEQUENCES: 10
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham LLP
CC STREET: 1185 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.24
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10579
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 42897-A-PCT/JPW/MAT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 278-0400
CC TELEFAX: (212) 391-0525
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 627 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 627 AA; 69878 MW; 2189782 CN;

Query Match 48.3%; Score 56; DB 4; Length 627;
Best Local Similarity 55.6%; Pred. No. 9.25e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 382 TMNPLSPLW 390
QY 3 AMDPTPLW 11
:|:|:|:|

RESULT 9
ID US-08-291-299-10 STANDARD: PRT: 627 AA.
XX
XX
XX
DT
XX
Sequence 10, Application US/08291299
XX
Sequence 10, Application US/08291299
XX
Patent No. 5766848
XX
GENERAL INFORMATION:
CC APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION
CC TITLE OF INVENTION: DNA ENCODING A HUMAN BETAINE/GABA
CC TITLE OF INVENTION: TRANSPORTER AND USES THEREOF

CC NUMBER OF SEQUENCES: 10
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.24
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/291,299
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 42897-A/JPW/TEP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 977-9550
CC TELEFAX: (212) 664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 627 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 627 AA; 69878 MW; 2189782 CN;

Query Match 48.3%; Score 56; DB 1; Length 627;
Best Local Similarity 55.6%; Pred. No. 9.25e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 382 TMNPLSPLW 390
QY 3 AMDPTPLW 11
:|:|:|:|

RESULT 10
ID US-09-100-851-1 STANDARD: PRT: 311 AA.
XX
XX
XX
DT
XX
Sequence 1, Application US/09100851
DE
XX
Sequence 1, Application US/09100851
CC
Patent No. 5911984
CC
GENERAL INFORMATION:
CC APPLICANT: Hillman, Jennifer L.
CC APPLICANT: Shah, Purvi
CC APPLICANT: Corley, Neil C.
CC TITLE OF INVENTION: HUMAN PEROXISOMAL THIOESTERASE
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/100,851

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jul 7 15:33:08 2000; MasPar time 4.47 Seconds.

Tabular output not generated. 158.218 Million cell updates/sec

Title: >US-09-300-612-1
Description: (1-15) from US09300612.pep
Perfect Score: 116
Sequence: 1 LKAMDPTPLWIKTE 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pif63
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.862; Variance 39.967; scale 0.672

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	91	78.4	237	2	A42013	3.78e-06
2	72	62.1	1213	2	A54063	2.70e-02
3	65	56.0	115	2	D71194	5.52e-01
4	64	55.2	225	2	C70045	8.37e-01
5	64	55.2	662	2	A29900	8.37e-01
6	64	55.2	1327	2	T09402	1.90e+00
7	62	53.4	271	2	S62187	8.37e-01
8	61	52.6	97	2	S59888	1.90e+00
9	59	50.9	102	1	J01890	2.85e+00
10	59	50.9	331	2	G55004	6.33e+00
11	59	50.9	564	1	VXHPWV	6.33e+00
12	58	50.0	184	2	B20807	9.37e+00
13	58	50.0	368	2	C72260	9.37e+00
14	58	50.0	492	2	T16659	9.37e+00
15	58	50.0	609	2	S65208	9.37e+00
16	58	50.0	1744	1	C4HU	9.37e+00
17	57	49.1	209	2	B40049	1.38e+01
18	57	49.1	355	2	A84138	1.38e+01
19	56	48.3	138	2	B69465	2.02e+01
20	56	48.3	162	2	C70829	2.02e+01
21	56	48.3	324	2	S46356	2.02e+01
22	56	48.3	328	2	T15555	2.02e+01
23	56	48.3	517	2	S69347	2.02e+01

24	55	47.4	132	2	JQ2270	hypothetical 15.2K pr	2.95e+01
25	55	47.4	263	2	S76994	hypothetical protein	2.95e+01
26	55	47.4	319	1	JC5644	acyl-CoA thiolesteras	2.95e+01
27	55	47.4	347	2	B70710	hypothetical protein	2.95e+01
28	55	47.4	357	1	TVHUL2	transforming protein	2.95e+01
29	55	47.4	425	2	C65059	hypothetical protein	2.95e+01
30	55	47.4	446	2	B70972	hypothetical protein	2.95e+01
31	55	47.4	453	2	T04646	aspartate transaminas	2.95e+01
32	55	47.4	522	1	S34190	sulfite reductase (NA	2.95e+01
33	55	47.4	552	1	A55760	cystathionine beta-sy	2.95e+01
34	55	47.4	594	2	B71893	exonuclease ABC chai	2.95e+01
35	55	47.4	594	1	E64622	Na+/Ca2+,K+-exchangin	2.95e+01
36	55	47.4	651	2	T03889	transketolase (EC 2.2	2.95e+01
37	55	47.4	680	1	XJBVTK	hypothetical protein	2.95e+01
38	55	47.4	695	2	E75099	Na+/Ca2+,K+-exchangin	2.95e+01
39	55	47.4	703	2	T03888	env polyprotein - sim	2.95e+01
40	55	47.4	864	1	VCLJG4	hydroxymethylglutaryl	2.95e+01
41	55	47.4	932	1	A31898	plasma protein xk - h	4.29e+01
42	54	46.6	41	2	PL0028	T-cell receptor alpha	4.29e+01
43	54	46.6	89	2	JQ0725	cyclin-dependent kina	4.29e+01
44	54	46.6	307	2	S50850	hypothetical protein	4.29e+01
45	54	46.6	912	2	D72644	hypothetical protein	4.29e+01

ALIGNMENTS

RESULT 1
ENTRY alpha-1-B-glycoprotein - North American opossum (fragments)
TITLE #formal_name Didelphis virginiana, Didelphis marsupialis
ORGANISM virginiana #common_name North American opossum
DATE 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change 31-Dec-1993
ACCESSIONS A42013
REFERENCE A42013
#authors Catanese, J.J.; Kress, L.F.
#journal Biochemistry (1992) 31:410-418
#title Isolation from opossum serum of a metalloproteinase inhibitor homologous to human alpha1B-glycoprotein.
#cross-references MUID:92118834
#accession A42013
##status preliminary
##molecule_type mRNA
##residues 1-237 ##label CAT
##cross-references GB:J05356
KEYWORDS glycoprotein
SUMMARY #length 237 #checksum 3610

Query Match 78.4%; Score 91; DB 2; Length 237;
Best Local Similarity 80.0%; Pred. No. 3.78e-06;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 LKAMDTPPLWIKTE 15
||||| ||||| |||||
QY 1 LKAMDTPPLWIKTE 15

RESULT 2
ENTRY A54063
TITLE TATA-binding protein-associated factor II - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 16-Apr-1999
ACCESSIONS A54063
REFERENCE A54063
#authors Verrijzer, C.P.; Yokomori, K.; Chen, J.L.; Tjian, R.
#journal Science (1994) 264:933-941
#title Drosophila TAF-II 150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA.
#cross-references MUID:94233377
#accession A54063
##status preliminary; not compared with conceptual translation

```

##molecule_type mRNA
##residues 1-1213 ##label VER
##cross-references GB:X79243; NID:G541664; PID:G541665
GENETICS
#gene FlyBase:Taf150
#length 1213 #molecular-weight 138533 #checksum 9849
SUMMARY
Query Match 62.18; Score 72; DB 2; Length 1213;
Best Local Similarity 53.3%; Pred. No. 2.70e-02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 620 LSAMDSPVLWIRLD 634
| | | | | | | | | |
QY 1 LKAMDTPPLWIKTE 15

RESULT 3
ENTRY D71194 #type complete
TITLE hypothetical protein PH1828 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
ACCESSION D71194
REFERENCE A71000
#authors Kavarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal DNA Res. (1998) 5:53-76
#title Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession D71194
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-115 ##label KAW
##cross-references GB:AP000007; NID:G3236134; PID:d1031890; PID:g3258264
##experimental_source strain OT3
##note this accession replaces an interim accession for a
sequence replaced by GenBank
GENETICS
#gene PH1828
SUMMARY
#length 115 #molecular-weight 12209 #checksum 6013
Query Match 56.08; Score 65; DB 2; Length 115;
Best Local Similarity 50.0%; Pred. No. 5.52e-01;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 88 TISPSPLWITS 99
| | | | | | | | | |
QY 3 AMDTPPLWIKT 14

RESULT 4
ENTRY C70045 #type complete
TITLE two-component response regulator [YvqB] homolog yvqA -
Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSION C70045
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

```

```

Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maeel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession C70045
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-225 ##label KUN
##cross-references GB:U09126; NID:G2635613;
PID:G2635798
##experimental_source strain 168
GENETICS
#gene yvqA
#classification #superfamily ompR protein; response regulator homology
#keywords phosphoprotein
#feature 5-113
52
#domain response regulator homology #label RHH\
#binding_site phosphate (Asp) (covalent) #status
predicted
SUMMARY
#length 225 #molecular-weight 26212 #checksum 1925
Query Match 55.28; Score 64; DB 2; Length 225;
Best Local Similarity 63.6%; Pred. No. 8.37e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 40 KKMTSPHLWI 50
| | | | | | | | | |
QY 2 KAMDTPPLWI 12

RESULT 5
ENTRY A29900 #type complete
TITLE fasciclin I precursor - American bird grasshopper
ORGANISM #formal_name Schistocerca americana #common_name American
bird grasshopper
DATE 22-Aug-1988 #sequence_revision 22-Aug-1988 #text_change
ACCESSION A29900
REFERENCE Zinn, K.; McAllister, L.; Goodman, C.S.
#authors Cell (1988) 53:577-587
#journal Sequence analysis and neuronal expression of fasciclin I in

```

```
grasshopper and Drosophila.
#cross-references MUID:88223351
#accession A29900
#status preliminary
##molecule_type mRNA
##residues 1-662 #label ZIN
##cross-references GB:M20544; GB:J03787; NID:g160846; PID:g160847
SUMMARY #length 662 #molecular-weight 75282 #checksum 9667

Query Match 55.2%; Score 64; DB 2; Length 662;
Best Local Similarity 58.3%; Pred. No. 8.37e-01;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 104 LSELDGNPLWI 115
QY 1 LKMDPTPLWI 12

RESULT 6
ENTRY T09402 #type complete
TITLE immunoglobulin-like protein IGSF1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999
ACCESSIONS T09402
REFERENCE Z16665
#authors Mazarella, R.; Penque, G.; Jones, J.; Jones, C.;
Schlessinger, D.
#journal Genomics (1998) 48:157-162
#title Cloning and expression of an immunoglobulin superfamily gene
(IGSF1) in Xq25.
#accession T09402
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-1327 #label MAZ
##cross-references EMBL:AF034198; NID:g2645889; PID:g2645890
GENETICS
#gene igsf1
#map_position Xq25
SUMMARY #length 1327 #molecular-weight 147971 #checksum 5419

Query Match 55.2%; Score 64; DB 2; Length 1327;
Best Local Similarity 58.3%; Pred. No. 8.37e-01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 25 MDPOPELWIESN 36
QY 4 MDPTPLWKTE 15

RESULT 7
ENTRY S62187 #type complete
TITLE thiosulfate sulfurtransferase (EC 2.8.1.1) - Azotobacter
vinelandii
ALTERNATE_NAMES rhodanese
ORGANISM #formal_name Azotobacter vinelandii
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change
26-Aug-1999
ACCESSIONS S62187; S62211
REFERENCE S62187
#authors Colnaghi, R.; Pagani, S.; Kennedy, C.; Drummond, M.
#journal Eur. J. Biochem. (1996) 236:240-248
#title Cloning, sequence analysis and overexpression of the
rhodanese gene of Azotobacter vinelandii.
#cross-references MUID:96184904
#accession S62187
##molecule_type DNA
##residues 1-271 #label COL
##cross-references EMBL:L42346; NID:gl069990; PID:gl069991
#accession S62211
##molecule_type protein
##residues 1-18 #label COW
GENETICS
```

```
rhda
#superfamily thiosulfate sulfurtransferase
sulfurtransferase
FEATURE 1-271
#product thiosulfate sulfurtransferase #status
experimental #label MAR
SUMMARY #length 271 #molecular-weight 29629 #checksum 3279

Query Match 53.4%; Score 62; DB 2; Length 271;
Best Local Similarity 53.8%; Pred. No. 1.90e+00;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 198 AMDPSRALRITD 210
QY 3 AMDPTPLWKTE 15

RESULT 8
ENTRY S59888 #type complete
TITLE C4 protein - tomato yellow leaf curl virus
ORGANISM #formal_name tomato yellow leaf curl virus
DATE 14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
20-Sep-1999
ACCESSIONS S59888
REFERENCE S58346
#authors Hong, Y.; Harrison, B.D.
#submission Submitted to the EMBL Data Library, February 1995
#description Nucleotide sequences from tomato leaf curl viruses from
different countries: evidence for three geographically
separate branches in evolution of the coat protein of
whitefly-transmitted geminiviruses.
#accession S59888
##status preliminary
##molecule_type DNA
##residues 1-97 #label HON
##cross-references EMBL:Z48182; NID:g944838; PIDN:CAA88232.1;
PID:g974214
CLASSIFICATION #superfamily tomato yellow leaf curl virus C4 protein
SUMMARY #length 97 #molecular-weight 11026 #checksum 3058

Query Match 52.6%; Score 61; DB 2; Length 97;
Best Local Similarity 46.7%; Pred. No. 2.85e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 42 LRAVQMSRPMWKTE 56
QY 1 LRAMDPTPLWKTE 15

RESULT 9
ENTRY JQ1890 #type complete
TITLE C4 protein - tomato yellow leaf curl virus (strain Australia)
ORGANISM #formal_name tomato yellow leaf curl virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
07-May-1999
ACCESSIONS JQ1890
REFERENCE JQ1885
#authors Dry, I.B.; Rigden, J.E.; Krake, L.R.; Mullineaux, P.M.;
Rezaian, M.A.
#journal J. Gen. Virol. (1993) 74:147-151
#title Nucleotide sequence and genome organization of tomato leaf
curl geminivirus.
#cross-references MUID:93139778
#accession JQ1890
##status translation not shown
##molecule_type DNA
##residues 1-102 #label DRY
##cross-references GB:S53251
CLASSIFICATION #superfamily tomato yellow leaf curl virus C4 protein
SUMMARY #length 102 #molecular-weight 11410 #checksum 1606

Query Match 50.9%; Score 59; DB 1; Length 102;
Best Local Similarity 40.0%; Pred. No. 6.33e+00;
```

```

Matches      6;  Conservative      6;  Mismatches      3;  Indels      0;  Gaps      0;

Db 44 LRAROTSSPIWRTTE 58
|:|:|:|:|:|:|
QY 1 LKAMDPTPLWKTE 15

RESULT 10
ENTRY G65004 #type complete
TITLE Div protein - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
14-Nov-1997
ACCESSIONS G65004
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession G65004
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-331 #label BLAT
##cross-references GB:AE000321; GB:U00096; NID:gl788659; PID:gl788661;
UMGP:B2321
##experimental_source strain K-12, substrain MG1655

GENETICS
#gene div
#summary #length 331 #molecular-weight 36668 #checksum 690

Query Match 50.9%; Score 59; DB 2; Length 331;
Best Local Similarity 50.0%; Pred. No. 6.33e+00;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 306 LQNMATRPALWI 317
|:|:|:|:|:|:|
QY 1 LKAMDPTPLWI 12

RESULT 11
ENTRY VHXPV #type complete
TITLE major structural nucleoprotein - Machupo virus
ALTERNATE_NAMES nucleocapsid protein
ORGANISM #formal_name Machupo virus
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
23-Jul-1999
ACCESSIONS SI8042
REFERENCE SI8042
#authors Griffiths, C.; Wilson, S.M.; Clegg, J.C.S.
#submission submitted to the EMBL Data Library, October 1991
#description Sequence of the nucleocapsid gene of Machupo virus: close
relationship with another South American pathogenic
arenavirus, Junin.
#accession SI8042
##molecule_type genomic RNA
##residues 1-564 #label GRI
##cross-references EMBL:X52616; NID:g60621; PIDN:CAA44486.1; PID:g60622

GENETICS
#map_position segment S
CLASSIFICATION #superfamily arenavirus major nucleoprotein
KEYWORDS nucleocapsid; nucleoprotein
SUMMARY #length 564 #molecular-weight 63299 #checksum 9818

Query Match 50.9%; Score 59; DB 1; Length 564;
Best Local Similarity 46.7%; Pred. No. 6.33e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 368 VKKLDPTNTLWDIE 382
|:|:|:|:|:|:|

```

```

|:|:|:|:|:|:|
QY 1 LKAMDPTPLWKTE 15

RESULT 12
ENTRY B20807 #type fragment
TITLE complement C4B - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-May-1988 #sequence_revision 21-May-1988 #text_change
29-Aug-1997
ACCESSIONS B20807
REFERENCE A90845
#authors Beit, K.T.; Carroll, M.C.; Porter, R.R.
#journal Cell (1984) 36:907-914
#title The structural basis of the multiple forms of human
complement component C4.
#cross-references MUID:84156544
#accession B20807
##molecule_type mRNA
##residues 1-184 #label BEL

GENETICS
#gene GDB:C4B
##cross-references GDB:119733; OMIM:120820
#map_position 6p21.3-6p21.3
CLASSIFICATION #superfamily alpha-2-macroglobulin
SUMMARY #length 184 #checksum 6928

Query Match 50.0%; Score 58; DB 2; Length 184;
Best Local Similarity 50.0%; Pred. No. 9.37e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 137 PMFOAPALWIET 148
|:|:|:|:|:|:|
QY 3 AMDPTPLWKI 14

RESULT 13
ENTRY C72260 #type complete
TITLE hypothetical protein - Thermotoga maritima (strain MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999
ACCESSIONS C72260
REFERENCE A72200
#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Uitterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession C72260
##status preliminary
##molecule_type DNA
##residues 1-368 #label ARN
##cross-references GB:AE001791; GB:AE000512; NID:g4981929; PID:g4981932;
TIGR:TM1370
##experimental_source strain MSB8

GENETICS
#gene TMI370
#summary #length 368 #molecular-weight 42191 #checksum 9660

Query Match 50.0%; Score 58; DB 2; Length 368;
Best Local Similarity 55.6%; Pred. No. 9.37e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 294 MDAIPSLWV 302
|:|:|:|:|:|:|
QY 4 MDPTPLWI 12

```

```
RESULT 14
ENTRY      T16659      #type complete
TITLE      hypothetical protein R02F2.2 - Caenorhabditis elegans
ORGANISM   #formal_name Caenorhabditis elegans
DATE       20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
ACCESSIONS T16659
REFERENCE   Z18555
            #authors   Pauley, A.
            #submission submitted to the EMBL Data Library, May 1994
            #description The sequence of C. elegans cosmid R02F2.
            #accession  T16659
            #status    Preliminary; translated from GB/EMBL/DBJ
            #molecule_type DNA
            #residues  1-492 ##label PAU
            ##cross-references EMBL:U00055; NID:g485143; PID:g485146;
            PIDN:AA50719.1; CESP:R02F2.2
            ##experimental_source strain Bristol N2
GENETICS
            #gene      CESP:R02F2.2
            #introns   36/2; 91/3; 154/2; 312/3; 327/2; 344/3; 415/3; 463/3
SUMMARY    #length 492 #molecular-weight 54302 #checksum 5641
            Query Match      50.0%; Score 58; DB 2; Length 492;
            Best Local Similarity 54.5%; Pred. No. 9.37e+00;
            Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 101 KRVDSTPSLWV 111
QY 2 KAMDTPPLWI 12

RESULT 15
ENTRY      S65208      #type complete
TITLE      probable membrane protein YPL189W - yeast (Saccharomyces
            cerevisiae)
ALTERNATE_NAMES hypothetical protein P2201
ORGANISM   #formal_name Saccharomyces cerevisiae
DATE       10-Dec-1994 #sequence_revision 31-May-1996 #text_change
ACCESSIONS S65208; S65201
REFERENCE   S65202
            #authors   Rieger, M.; Mueller-Auer, S.; Schaefer, M.
            #submission submitted to the Protein Sequence Database, May 1996
            #accession  S65208
            #molecule_type DNA
            #residues  1-609 ##label RTE
            ##cross-references EMBL:Z73545; NID:gl370394; PID:e246916; PID:gl370395;
            MIPS:YPL189W
            ##experimental_source strain S288C (AB972)
REFERENCE   S65183
            #authors   Benes, V.; Rechmann, S.; Nentwich, U.; Voss, H.; Ansorge, W.
            #submission submitted to the Protein Sequence Database, May 1996
            #accession  S65201
            #molecule_type DNA
            #residues  177-609 ##label BEN
            ##cross-references EMBL:Z73545; MIPS:YPL189W
            ##experimental_source strain S288C (AB972)
GENETICS
            #map_position 16L
KEYWORDS   transmembrane protein
FEATURE
79-95      #domain transmembrane #status predicted #label TM1\
136-152    #domain transmembrane #status predicted #label TM2\
164-180    #domain transmembrane #status predicted #label TM3\
201-217    #domain transmembrane #status predicted #label TM4\
328-344    #domain transmembrane #status predicted #label TM5\
376-392    #domain transmembrane #status predicted #label TM6\
406-422    #domain transmembrane #status predicted #label TM7\
498-514    #domain transmembrane #status predicted #label TM8\
534-550    #domain transmembrane #status predicted #label TM9\
```

```
578-594    #domain transmembrane #status predicted #label TM10
SUMMARY    #length 609 #molecular-weight 71288 #checksum 4342
            Query Match      50.0%; Score 58; DB 2; Length 609;
            Best Local Similarity 54.5%; Pred. No. 9.37e+00;
            Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 59 LSSNSPSPPLW 69
QY 1 LKAMDTPPLW 11
            Search completed: Fri Jul 7 15:33:18 2000
            Job time : 10 secs.
```

WPSRCH_PP

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm

Run On: Fri Jul 7 15:33:35 2000; MasPar time 7.49 Seconds

Tabular output not generated.
138.914 Million cell updates/sec

Title: >US-09-300-612-1
Description: (1-15) from US09300612.pep
Perfect Score: 116
Sequence: 1 LKAMPPTPLWKTE 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.585; Variance 35.674; scale 0.745

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	65	56.0	115	1	059492	115AA LONG HYPOTHETICAL
2	64	55.2	216	2	032304	YIRD.
3	64	55.2	225	2	032192	YVQA PROTEIN.
4	64	55.2	836	11	P97359	TAFI95.
5	64	55.2	1189	4	043604	TATA BINDING PROTEIN A
6	64	55.2	1199	4	043487	COFACTOR OF INITIATOR
7	64	55.2	1199	4	060668	TBP-ASSOCIATED FACTOR
8	64	55.2	1327	4	015070	KIAA0364.
9	61	52.6	97	14	Q88560	V1, V2, C1, C2, C3 AND
10	61	52.6	258	5	Q17984	R06B9.2 PROTEIN.
11	60	51.7	322	10	Q32WF1	THYMIDYLATE SYNTHASE.
12	59	50.9	354	5	076637	T08E11.6 PROTEIN.
13	58	50.0	100	4	Q14033	COMPLEMENT COMPONENT C
14	58	50.0	368	2	Q9X190	HYPOTHETICAL 42.2 KD P
15	58	50.0	381	4	Q14835	COMPLEMENT COMPONENT C
16	58	50.0	395	2	Q47207	FASG PRECURSOR.
17	58	50.0	609	3	Q08929	CHROMOSOME XVI READING
18	58	50.0	766	10	Q64790	TLF9.21.
19	58	50.0	925	10	Q64789	TLF9.20.
20	58	50.0	1699	4	Q33160	COMPLEMENT C4B PRECURS

21	58	50.0	1744	4	Q13906	COMPLEMENT COMPONENT C
22	57	49.1	97	14	Q41338	HYPOTHETICAL 11.2 KD P
23	57	49.1	97	14	Q920B3	C1 AND C4 GENES, CLONE
24	56	48.3	138	1	Q28551	TRANSCRIPTIONAL REGULA
25	56	48.3	162	2	Q53756	HYPOTHETICAL 17.6 KD P
26	56	48.3	324	2	Q52209	PUP1 AND PUPR GENES.
27	56	48.3	403	5	P91736	LM-AMID AND MW-AMID-CO
28	56	48.3	506	10	Q9XF4	PREPROCARDOSIN B PRECU
29	56	48.3	873	13	Q98949	AXL-RELATED RECEPTOR
30	56	48.3	1794	14	P87515	NONSTRUCTURAL POLYPROT
31	55	47.4	132	2	Q44307	RIBULOSE 1,5-BISPHOSPH
32	55	47.4	263	2	Q55923	HYPOTHETICAL 28.8 KD P
33	55	47.4	319	4	Q14734	HIV-NEF ASSOCIATED ACY
34	55	47.4	319	4	Q15261	THIOESTERASE II.
35	55	47.4	347	2	P71844	HYPOTHETICAL 37.7 KD P
36	55	47.4	446	2	Q50400	HYPOTHETICAL 48.8 KD P
37	55	47.4	453	10	Q49392	ASPARTATE AMINOTRANSFE
38	55	47.4	457	10	Q22618	ASPARTATE AMINOTRANSFE
39	55	47.4	594	2	Q92L21	EXCINUCLEASE ABC SUBUN
40	55	47.4	651	5	Q16242	C13D9.8 PROTEIN.
41	55	47.4	703	5	Q16241	C13D9.7 PROTEIN.
42	55	47.4	792	5	Q18866	COSMID C55C3.
43	55	47.4	1449	14	Q65974	ORF 1.
44	55	47.4	1454	5	Q10463	T24H7.5 PROTEIN.
45	55	47.4	1569	10	Q22997	F6P23.4 PROTEIN.

ALIGNMENTS

RESULT 1
ID 059492 PRELIMINARY; PRT; 115 AA.
AC 059492;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-JAN-1999 (TEMBLrel. 09, Last annotation update)
DE 115AA LONG HYPOTHETICAL PROTEIN.
GN PH1828.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res 5:55-76(1998).
DR EMBL; AP000007; BAA30947.1; -.
SQ SEQUENCE 115 AA; 12209 MW; 1846C3BA CRC32;

Query Match 56.0%; Score 65; DB 1; Length 115;
Best Local Similarity 50.0%; Pred.No. 2.40e-01;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 88 TISFSPPLWITS 99
:::|||||:
QY 3 AMDPTPLWKTK 14

RESULT 2
ID 032304 PRELIMINARY; PRT; 216 AA.
AC 032304;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE YIRD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;

```

Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=168;
RC WIPAT A., BRIGNELL C.S., GUY J.B., ROSE M., EMMERSON P.T.,
RA HARWOOD C.R.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR ENBL; Z99120; CAB15291.1; -.
DR EMBL; AJ223978; CAAL1752.1; -.
DR DR HSP; P03025; IODD.
DR PFAM; PF00072; response_reg; 1.
DR PFAM; PF00486; trans_reg_C; 1.
SQ SEQUENCE 225 AA; 26212 MW; FDI1BD0D CRC32;

Query Match 55.2%; Score 64; DB 2; Length 225;
Best Local Similarity 63.8%; Pred. No. 3.82e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0;

Db 40 KMTPTSPHLWI 50
| | | | |
QY 2 KAMDTPPLWI 12

RESULT 4
ID P97359 PRELIMINARY; PRT; 836 AA.
AC P97359;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
TA F1195.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA HEIX J., ZOMERDIJK J.C.M.B., RAVANPAY A., TJIAN R., GRUMMT I.
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR ENBL; Y09974; CAA71093.1; -.
SQ SEQUENCE 836 AA; 92037 MW; A640D245 CRC32;

Query Match 55.2%; Score 64; DB 11; Length 836;
Best Local Similarity 50.0%; Pred. No. 3.82e-01;
Matches 6; Conservative 3; Mismatches 3; Indels 0;

Db 612 LMELSPTRPLW 623
| | | | |
QY 1 LKAMDTPPLWI 12

RESULT 5
ID 043604 PRELIMINARY; PRT; 1189 AA.
AC 043604;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE TATA BINDING PROTEIN ASSOCIATED FACTOR.
GN TAF1150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MARTINEZ E., GE H., TAO Y., YUAN C.-X., ROEDER R.G.;
RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF040701; AAC68502.1; -.
SQ SEQUENCE 1189 AA; 135833 MW; 180759F2 CRC32;

Query Match 55.2%; Score 64; DB 4; Length 1189;
Best Local Similarity 50.0%; Pred. No. 3.82e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 1;

Db 606 LSAMDADSPLLWIRID 621
| | | | |

```

QY 1 LKAMDPTPL-WIKTE 15

RESULT 6
ID O43487 PRELIMINARY; PRT; 1199 AA.
AC O43487;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE COFACTOR OF INITIATOR FUNCTION.
GN C1F150
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98078679.
RA KAUFMANN J., AHRENS K., KOOP R., SMALE S.T., MULLER R.;
RT "C1F150, a human cofactor for transcription factor IID-dependent
RT initiator function."
RL Mol. Cell. Biol. 18:233-239(1998).
DR EMBL; AF026445; AAC02966.1; -.
SQ SEQUENCE 1199 AA; 136993 MW; A6363760 CRC32;

Query Match 55.2%; Score 64; DB 4; Length 1199;
Best Local Similarity 50.0%; Pred. No. 3.82e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 616 LSAMDADSPLLWIRID 631

QY 1 LKAMDPTPL-WIKTE 15

RESULT 7
ID O60668 PRELIMINARY; PRT; 1199 AA.
AC O60668;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TReMBLrel. 07, Last annotation update)
DE TBP-ASSOCIATED FACTOR TAFII150.
GN TAFII150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA GUERMAH M., ROEDER R.G.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF057694; AAC13540.1; -.
SQ SEQUENCE 1199 AA; 136985 MW; 98118DA8 CRC32;

Query Match 55.2%; Score 64; DB 4; Length 1199;
Best Local Similarity 50.0%; Pred. No. 3.82e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 616 LSAMDADSPLLWIRID 631

QY 1 LKAMDPTPL-WIKTE 15

RESULT 8
ID O15070 PRELIMINARY; PRT; 1327 AA.
AC O15070;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE KIAA0364.
GN KIAA0364.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.

TISSUE-BRAIN;
RC MEDLINE; 97349984.
RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002362; BAA20819.1; -.
DR HSP; P43626; INKR.
DR PFAM; PF00047; Ig; 10.
SQ SEQUENCE 1327 AA; 147971 MW; 0A68A862 CRC32;

Query Match 55.2%; Score 64; DB 4; Length 1327;
Best Local Similarity 58.3%; Pred. No. 3.82e-01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 25 MDQPELWIESN 36

QY 4 MDPTPLWIKTE 15

RESULT 9
ID Q88560 PRELIMINARY; PRT; 97 AA.
AC Q88560;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE V1, V2, C1, C2, C3 AND C4 GENES.
GN C4.
OS Tomato yellow leaf curl virus (TYLCV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INDIAN TOMATO LEAF CURL VIRUS (ITMLCV);
RX MEDLINE; 95363424.
RA HONG Y., HARRISON B.D.;
RT "Nucleotide sequences from tomato leaf curl viruses from different
RT countries: evidence for three geographically separate branches in
RT evolution of the coat protein of whitefly-transmitted geminiviruses."
RL J. Gen. Virol. 76:2043-2049(1995).
DR EMBL; Z48182; CAA88232.1; -.
DR PFAM; PF01492; Gemini_C4; 1.
SQ SEQUENCE 97 AA; 11026 MW; 62CE9CD2 CRC32;

Query Match 52.6%; Score 61; DB 14; Length 97;
Best Local Similarity 46.7%; Pred. No. 1.49e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 42 LRAYOMSRPMWKTE 56

QY 1 LKAMDPTPLWIKTE 15

RESULT 10
ID O17984 PRELIMINARY; PRT; 258 AA.
AC O17984;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE R06B9.2 PROTEIN.
GN R06B9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA BAYNES C.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMDALON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RN Nature 368:32-38(1994).
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA JOHNSON D., WAMSLEY P., KRAMER J., ELLIOTT G.;
 RT "The sequence of C. elegans cosmid T08E11.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 DR EMBL: AF077546; AAC26314.1; -
 DR PFAM: PF00646; F-box; 1
 SQ SEQUENCE 258 AA; 29059 MW; CFF322C5 CRC32;

Query Match 52.6%; Score 61; DB 5; Length 258;
 Best Local Similarity 50.0%; Pred. No. 1.49e+00;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 79 KHLDTAVPLWRKTD 92
 QY 2 KAMDPTPLWIKTE 15
 : : | : | | : | :
 : : | : | | : | :
 : : | : | | : | :

RESULT 11
 ID Q92WEL PRELIMINARY; PRT; 322 AA.
 AC Q92WEL;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-NOV-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE THYMIDYLATE SYNTHASE.
 GN THYA.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Oryza.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KANJO N., INOKUCHI H.;
 RT "Nucleotide sequence of a cDNA clone encoding thymidylate synthase
 RT from rice (Accession No. AB023402).";
 RL Plant Physiol. 120:634-634(1999).
 DR EMBL: AB023402; BAA74947.1; -
 DR HSP: P00469; ITSM.
 SQ SEQUENCE 322 AA; 36506 MW; AA7097E2 CRC32;

Query Match 51.7%; Score 60; DB 10; Length 322;
 Best Local Similarity 35.7%; Pred. No. 2.33e+00;
 Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 280 RPMHPLKWLNPD 293
 QY 2 KAMDPTPLWIKTE 15
 : : | : | | : | :
 : : | : | | : | :

RESULT 12
 ID O76637 PRELIMINARY; PRT; 354 AA.
 AC O76637;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE T08E11.6 PROTEIN.
 GN T08E11.6
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMDALON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RN Nature 368:32-38(1994).
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA JOHNSON D., WAMSLEY P., KRAMER J., ELLIOTT G.;
 RT "The sequence of C. elegans cosmid T08E11.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 DR EMBL: AF077546; AAC26314.1; -
 DR PFAM: PF00646; F-box; 1
 SQ SEQUENCE 354 AA; 40564 MW; 21878A8C CRC32;

Query Match 50.9%; Score 59; DB 5; Length 354;
 Best Local Similarity 50.0%; Pred. No. 3.61e+00;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 45 LRHMPTTQFWF 56
 QY 1 LKAMDPTPLWI 12
 : : | : | | : | :
 : : | : | | : | :

RESULT 13
 ID Q14033 PRELIMINARY; PRT; 100 AA.
 AC Q14033;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE COMPLEMENT COMPONENT C4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 83117835.
 RA CARROLL M.C., PORTER R.R.;
 RT "Cloning of a human complement component C4 gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:264-267(1983).
 DR EMBL: Y00502; CAA23760.1; -
 DR PFAM: PF00207; A2M; 1.
 SQ SEQUENCE 100 AA; 10629 MW; 4589044E CRC32;

Query Match 50.0%; Score 58; DB 4; Length 100;
 Best Local Similarity 50.0%; Pred. No. 5.58e+00;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 59 PMQPAPALWIET 70
 QY 3 AMDPTPLWIKT 14
 : : | : | | : | :
 : : | : | | : | :

RESULT 14
 ID Q9X190 PRELIMINARY; PRT; 368 AA.
 AC Q9X190;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE HYPOTHETICAL 42.2 KD PROTEIN.
 GN TM1370.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.

```

RN SEQUENCE FROM N.A.
RP MEDLINE; 99287316.
RX NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RT "Evidence for lateral gene transfer between Archaea and bacteria from
RL genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
RN [12]
RP SEQUENCE FROM N.A.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE001791; AAD36441.1; -
KW Hypothetical protein.
SQ SEQUENCE 368 AA; 42191 MW; D6E055B8 CRC32;

```

```

Query Match      50.0%; Score 58; DB 2; Length 368;
Best Local Similarity 55.6%; Pred. No. 5.58e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 294 MDAIPSLWV 302
||:|:|:|
QY 4 MDPTPLWI 12

```

```

RESULT 15
ID Q14835 PRELIMINARY; PRT; 381 AA.
AC Q14835;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE COMPLEMENT COMPONENT C4B (FRAGMENT).
GN C4B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 84156544.
RA BELT K.T., CARROLL M.C., PORTER R.R.;
RT "The structural basis of the multiple forms of human complement
RT component C4.";
RL Cell 36:907-914(1984).
DR EMBL; K02404; AAA59651.1; -.
DR HSSP; P01024; IC3D.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR PFAM; PF00207; A2M; 1.
FT NON_TER 381
SQ SEQUENCE 381 AA; 40762 MW; 7A5E863A CRC32;

```

```

Query Match      50.0%; Score 58; DB 4; Length 381;
Best Local Similarity 50.0%; Pred. No. 5.58e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 298 PMEQAPALWET 309
:|:|:|:|
QY 3 AMDPTPLWIKT 14

```

```

Search completed: Fri Jul 7 15:33:53 2000
Job time : 18 secs.

```

WISREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jul 7 15:32:41 2000; Maspar time 3.01 Seconds
Tabular output not generated. 151.654 Million cell updates/sec

Title: >US-09-300-612-1
Description: (1-15) from US09300612.pep
Sequence: 1 LRAMPDTPPLWKTE 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 27.363; Variance 36.961; scale 0.740

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	72	62.1	1213	1 T2D2_DROME	TRANSCRIPTION INITIATI	6.27e-03
2	64	55.2	662	1 FAS1_SCHAM	FASCICLIN I PRECURSOR	2.59e-01
3	62	53.4	271	1 THR_AZOVI	THIOSULFATE SULFURTRAN	6.30e-01
4	61	52.6	1220	1 PCL1_BRARE	PATCHED PROTEIN HOMOLO	9.77e-01
5	59	50.9	102	1 YC4_TYLCA	HYPOTHETICAL 11.4 KD P	2.31e+00
6	59	50.9	331	1 DIV_ECOLI	DIV PROTEIN	2.31e+00
7	59	50.9	453	1 MTE1_RAT	ACYL COENZYME A THIOES	2.31e+00
8	59	50.9	564	1 NCAP_MACHU	NUCLEOCAPSID PROTEIN (2.31e+00
9	59	50.9	1227	1 LAF4_HUMAN	LAF-4 PROTEIN (LYMPHOI	2.31e+00
10	58	50.0	351	1 Y4VJ_RHISN	HYPOTHETICAL 39.2 KD P	3.53e+00
11	58	50.0	492	1 YNG2_CAEEL	HYPOTHETICAL 54.3 KD P	3.53e+00
12	58	50.0	1305	1 RPL1_AHSV9	RNA-DIRECTED RNA POLYM	3.53e+00
13	58	50.0	1741	1 CO4_HUMAN	COMPLEMENT C4 PRECURSO	3.53e+00
14	57	49.1	209	1 YSCK_YERPS	YOP PROTEINS TRANSLOCA	5.36e+00
15	57	49.1	209	1 YSCK_YEREN	YOP PROTEINS TRANSLOCA	5.36e+00
16	57	48.1	355	1 RFE_RAEIN	PUTATIVE UNDECAPENYL-	5.36e+00
17	56	48.3	192	1 Y4PG_RHISN	HYPOTHETICAL 21.1 KD P	8.10e+00
18	56	48.3	328	1 SRG1_CAEEL	SRG-1 PROTEIN	8.10e+00
19	56	48.3	517	1 CPN1_RANCA	CYTOCHROME P450 11B PR	8.10e+00
20	55	47.4	357	1 MYCM_HUMAN	L-MYC-2 PROTEIN	1.22e+01
21	55	47.4	419	1 YCEI_MOUSE	CYTOSOLIC ACYL COENZYM	1.22e+01
22	55	47.4	425	1 YCEI_ECOLI	HYPOTHETICAL 46.8 KD P	1.22e+01
23	55	47.4	446	1 YX71_MYCTU	HYPOTHETICAL 48.8 KD P	1.22e+01

24	55	47.4	453	1 AATM_ARATH	ASPARTATE AMINOTRANSFE	1.22e+01
25	55	47.4	522	1 CYSJ_THIRO	SULFITE REDUCTASE [NAD	1.22e+01
26	55	47.4	550	1 CBS_HUMAN	CYSTATHIONINE BETA-SYN	1.22e+01
27	55	47.4	594	1 UVR_C_HELPY	EXCINUCLEASE ABC SUBUN	1.22e+01
28	55	47.4	679	1 TKT1_YEAST	TRANSKETOLASE 1 (EC 2.	1.22e+01
29	55	47.4	865	1 ENV_SIVAT	ENVELOPE POLYPROTEIN G	1.22e+01
30	55	47.4	932	1 HMDH_STRPU	3-HYDROXY-3-METHYLGLUT	1.22e+01
31	54	46.6	41	1 ALBG_HORSE	ALPHA-1B-GLYCOPROTEIN	1.82e+01
32	54	46.6	112	1 MTRB_METKA	TETRAHYDROMETHANOPTERI	1.82e+01
33	54	46.6	171	1 TRAV_ECOLI	TRAV PROTEIN PRECURSOR	1.82e+01
34	54	46.6	307	1 CD5R_MOUSE	CYCLIN-DEPENDENT KINAS	1.82e+01
35	54	46.6	307	1 CD5R_HUMAN	CYCLIN-DEPENDENT KINAS	1.82e+01
36	54	46.6	307	1 CD5R_BOVIN	CYCLIN-DEPENDENT KINAS	1.82e+01
37	54	46.6	396	1 TCR3_ECOLI	TETRACYCLINE RESISTANC	1.82e+01
38	54	46.6	564	1 NCAP_JUNIN	NUCLEOCAPSID PROTEIN (1.82e+01
39	54	46.6	777	1 YASB_SCHPO	HYPOTHETICAL 89.6 KD P	1.82e+01
40	54	46.6	986	1 EPA4_HUMAN	EPHRIN TYPE-A RECEPTOR	1.82e+01
41	54	46.6	986	1 EPA4_CHICK	EPHRIN TYPE-A RECEPTOR	1.82e+01
42	54	46.6	1278	1 DHBF_BAGSU	PROBABLE SERINE ACTIVA	1.82e+01
43	53	45.7	73	1 I73_ASPE7	EARLY PROTEIN I73R	2.70e+01
44	53	45.7	558	1 NCAP_LICVW	NUCLEOCAPSID PROTEIN (2.70e+01
45	53	45.7	883	1 HSS2_MOUSE	HEPARIN SULFATE N-DEAC	2.70e+01

ALIGNMENTS

RESULT 1
ID T2D2_DROME STANDARD; PRT; 1213 AA.
AC Q24325;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TRANSCRIPTION INITIATION FACTOR TFIID 150 KD SUBUNIT (TAFII-150)
DE (TAFII150).
GN TAFI50.

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=EMBRYO;
RX MEDLINE; 94233377.
RA Verrijzer C.P., Yokomori K., Chen J.-L., Tjian R.;
RT "Drosophila TAFII150: similarity to yeast gene TSM-1 and specific
RT binding to core promoter DNA.";
RL Science 264:933-941(1994).
CC -1- FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID
CC (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF RNA
CC POLYMERASE TRANSCRIPTION. TAFII-150 IS AN ESSENTIAL SUBUNIT WHICH
CC INTERACTS DIRECTLY WITH TBP AND TAFII-250, AND BINDS TO CORE
CC PROMOTOR DNA.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TO YEAST TAFII-150 (TSM1).
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; X79243; CAA55830.1; -;
CC TRANSFAC; T02120; -;
CC FLYBASE; FBgn0011836; Tafi50.
CC Transcription regulation; Nuclear protein.
KW DOMAIN 845 1213 BINDS TO TBP AND TAFII-250.
FT DOMAIN 1138 1183 HIGHLY CHARGED.
SQ SEQUENCE 1213 AA; 138533 MW; 72A5B473E26FD064 CRC64;

Query Match 62.18; Score 72; DB 1; Length 1213;
Best Local Similarity 53.3%; Pred. No. 6.27e-03;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 620 LSAMDSPVLWIRLD 634
| | | | | | | | | |
QY 1 LKAMDPTPLWKTE 15

RESULT 2
ID FAS1_SCHAM STANDARD; PRT; 662 AA.
AC P10675;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE FASCICLIN I PRECURSOR (FAS I) (FCN).
FAS1.
GN Schistocerca americana (American grasshopper).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridoidea; Acridoidea; Acrididae; Cyrtacanthacridinae;
OC Schistocerca.
RN [1]
CC SEQUENCE FROM N.A.
CC MEDLINE; 88223351.
RA Zinn K., McAllister L., Goodman C.;
RT "Sequence analysis and neuronal expression of fasciclin I in
RT grasshopper and Drosophila";
RL Cell 53:577-587(1988).
RN [2]
CC SEQUENCE OF 25-42.
RX MEDLINE; 88276943.
RA Snow P.M., Zinn K., Harrelson A.L., McAllister L., Schilling J.,
RA Bastiani M.J., Makk G., Goodman C.S.;
RT "Characterization and cloning of fasciclin I and fasciclin II
RT glycoproteins in the grasshopper";
RL Proc. Natl. Acad. Sci. U.S.A. 85:5291-5295(1988).
CC -1- FUNCTION: NEURAL CELL ADHESION MOLECULE.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: EXPRESSED ON DIFFERENT SUBSETS OF AXON BUNDLES
CC (FASCICLES) IN INSECT EMBRYOS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M20544; AAA29809.1; -
PIR; A31817; A31817.
DR PIR; A29900; A29900.
KW Cell adhesion; Glycoprotein; Repeat; Signal; GPI-anchor.
FT SIGNAL 1 24 FASCICLIN I.
FT CHAIN 25 662
FT REPEAT 37 185 1.
FT REPEAT 194 334 2.
FT REPEAT 339 483 3.
FT REPEAT 484 637 4.
FT REPEAT 54 54 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT CARBOHYD 437 437 POTENTIAL.
FT CARBOHYD 448 448 POTENTIAL.
FT CARBOHYD 488 488 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 662 AA; 75282 MW; 1E648D139A16B816 CRC64;

Query Match 55.28; Score 64; DB 1; Length 662;
Best Local Similarity 58.3%; Pred. No. 2.59e-01;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

DB 104 LSELDCNPPLWI 115

QY 1 LKAMDPTPLWI 12
| | | | | | | | | |

RESULT 3
ID THTR_AZOVI STANDARD; PRT; 271 AA.
AC P52197;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE THIOSULFATE SULFOTRANSFERASE (EC 2.8.1.1) (RHODANESE-LIKE PROTEIN).
GN RHD.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae;
OC Azotobacter.
RN [1]
CC SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
CC STRAIN-OP / UW136;
CC MEDLINE; 96184904.
RA Colnaghi R., Pagani S., Kennedy C., Drummond M.;
RT "Cloning, sequence analysis and overexpression of the rhodanese gene
RT of Azotobacter vinelandii";
RL Eur. J. Biochem. 236:240-248(1996).
CC -1- CATALYTIC ACTIVITY: THIOSULFATE + CYANIDE - SULFITE + THIOCYANATE.
CC -1- SIMILARITY: BELONGS TO THE RHODANESE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L42346; AAB03239.1; -
DR PROSITE; PS00380; RHODANESE_1; 1.
DR PROSITE; PS00683; RHODANESE_2; 1.
DR PFAM; PF00581; Rhodanese; 2.
KW Transferase.
FT ACT_SITE 230 230 BY SIMILARITY.
FT ACT_SITE 230 230
SQ SEQUENCE 271 AA; 29629 MW; 47E573D2D34EA77C CRC64;

Query Match 53.4%; Score 62; DB 1; Length 271;
Best Local Similarity 53.8%; Pred. No. 6.30e-01;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 198 AMDPSRALRIKD 210
| | | | | | | | | |

QY 3 AMDPTPLWKTE 15

RESULT 4
ID PTCL_BRARE STANDARD; PRT; 1220 AA.
AC Q98864;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PATCHED PROTEIN HOMOLOG 1 (PATCHED 1) (PTC1).
GN PTC1.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinidae; Rasbora; Danio.
RN [1]
CC SEQUENCE FROM N.A.
CC TISSUE-EMBRYO;
CC MEDLINE; 9639744.
RX Concordet J.-P., Lewis K.E., Moore J.W., Goodrich L.V., Johnson R.L.,
RA Scott M.P., Ingham P.W.;
RT "Spatial regulation of the zebrafish patched homologue reflects the
RT roles of sonic hedgehog and protein kinase A in neural tube and somite
RT patterning";
RL Development 122:2835-2846(1996).

```

CC -1- FUNCTION: ACTS AS A RECEPTOR FOR SONIC HEDGEHOG (SHH), INDIAN
CC HEDGEHOG (HHH) AND DESERT HEDGEHOG (DHH). ASSOCIATES WITH THE
CC SMOOTHENED PROTEIN (SMO) TO TRANSDUCE THE HEDGEHOG'S PROTEINS
CC SIGNAL (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: DETECTED IN EMBRYONIC PRESOMITIC MESODERM,
CC NEUROECTODERM, TISSUE SURROUNDING THE NOTOCHORD, VENTRAL NEURAL
CC TUBE.
CC -1- DEVELOPMENTAL STAGE: AT ALL STAGES, EXPRESSION CORRESPONDS TO THE
CC LOCALIZATION OF SHH. FIRST DETECTED DURING GASTRULATION. BY 36
CC HOURS, PTC1 APPEARS IN THE FIRST BRANCHIAL ARCH AND THE POSTERIOR
CC MESENCHYME OF THE FIN BUD; BY 48 HOURS, IN THE HINDBRAIN AND
CC FOREBUD.
CC -1- INDUCTION: ACTIVATED BY SONIC HEDGEHOG.
CC -1- PTM: GLYCOSYLATION IS NECESSARY FOR SHH BINDING (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PATCHED FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98883; CAA67386.1; -
DR ZFIN; ZDB-GENE-980526-44; PTC1.
KW Receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 105 POTENTIAL.
FT DOMAIN 106 419 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 420 440 POTENTIAL.
FT DOMAIN 441 449 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 450 470 POTENTIAL.
FT DOMAIN 471 484 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 485 505 POTENTIAL.
FT DOMAIN 506 528 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 529 549 POTENTIAL.
FT DOMAIN 550 558 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 559 579 POTENTIAL.
FT DOMAIN 580 739 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 740 760 POTENTIAL.
FT DOMAIN 761 1016 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1017 1037 POTENTIAL.
FT DOMAIN 1038 1044 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1045 1065 POTENTIAL.
FT DOMAIN 1066 1072 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1073 1093 POTENTIAL.
FT DOMAIN 1094 1110 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1111 1131 POTENTIAL.
FT DOMAIN 1132 1143 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1144 1164 POTENTIAL.
FT DOMAIN 1165 1220 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 397 397 POTENTIAL.
FT CARBOHYD 865 865 POTENTIAL.
FT CARBOHYD 888 888 POTENTIAL.
SQ SEQUENCE 1220 AA; 135544 MW; D10A9D04115F532D CRC64;

Query Match 52.6%; Score 61; DB 1; Length 1220;
Best Local Similarity 46.2%; Pred. No. 9.77e-01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 53 KAVGQKAPLMWIRA 65
   ||: |||||:
QY 2 KAMDPTPLWIKT 14

RESULT 5
ID YC4_TYLC4 STANDARD; PRT; 102 AA.
AC P36283;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)

```


RC STRAIN=K12;
RX MEDLINE; 90036695.
RA Schoenlein P.V., Roa B.B., Winkler M.E.;
RT "Divergent transcription of pdxB and homology between the pdxB and
RT serA gene products in *Escherichia coli* K-12.";
RL J. Bacteriol. 171:6084-6092(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000321; AAC75381.1; -
DR EMBL; D90863; CAB22095.1; -
DR EMBL; U76961; AAB36529.1; -
DR EMBL; M29962; CAB25573.1; -
DR EMBL; M15541; -; NOT_ANNOTATED_CDS.
DR ECGENE; EG10229; div.
SQ SEQUENCE 331 AA; 36668 MW; 9201E5BF4B9D27FB CRC64;

Query Match 50.9%; Score 59; DB 1; Length 331;
Best Local Similarity 50.0%; Pred. No. 2.31e+00;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 306 LQNMATRPALWI 317
| | | | |
QY 1 LKAMDPTPLWI 12

RESULT 7
ID MTEL RAT STANDARD; PRT; 453 AA.
AC O55171; O88268;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACYL COENZYME A THIOESTER HYDROLASE, MITOCHONDRIAL PRECURSOR
DE (EC 3.1.2.2) (VERY-LONG-CHAIN ACYL-COA THIOESTERASE) (MTE-I).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 147-166 AND 168-178.
RC STRAIN=Sprague-Dawley;
RX MEDLINE; 98109736.
RA Svensson L.T., Engberg S.T., Aoyama T., Usuda N., Alexson S.E.H.,
RA Hashimoto T.;
RT "Molecular cloning and characterization of a mitochondrial peroxisome
RT proliferator-induced acyl-CoA thioesterase from rat liver.";
RL Biochem. J. 329:601-608(1998).
RN [2]
RP SEQUENCE OF 1-197 FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 98369597.
RA Yamada J., Suga K., Furihata T., Kitahara M., Watanabe T.,
RA Hosokawa M., Satoh T.;
RT "cDNA cloning and genomic organization of peroxisome proliferator-
RT inducible long-chain acyl-CoA hydrolase from rat liver cytosol.";
RL Biochem. Biophys. Res. Commun. 248:608-612(1998).
RN [3]
RP CHARACTERIZATION.
RC TISSUE=LIVER;
RX MEDLINE; 95263569.
RA Svensson L.T., Alexson S.E., Hiltunen J.K.;
RT "Very long chain and long chain acyl-CoA thioesterases in rat liver
RT mitochondria. Identification, purification, characterization, and
RT induction by peroxisome proliferators.";
RL J. Biol. Chem. 270:12177-12183(1995).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN LIPID METABOLISM. MOST ACTIVE
CC ON SUBSTRATES WITH CHAIN LENGTHS RANGING FROM C14-C20. IT HAS A PH
CC OPTIMUM BETWEEN 8 AND 9.

CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN HEART AND BROWN
CC FAT. STRONGLY INDUCED IN LIVER, AND WEAKLY IN KIDNEY, IN
CC PEROXISOME PROLIFERATOR TREATED RAT.
CC -1- PTM: THE N-TERMINAL IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE ACYL COENZYME A THIOESTER HYDROLASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y09333; CAA70513.1; -
DR EMBL; AB010429; BAA32539.1; -
KW Hydrolase; Serine esterase; Mitochondrion; Transit peptide.
FT TRANSIT 1 42 MITOCHONDRION (POTENTIAL).
FT CHAIN 43 453 ACYL COENZYME A THIOESTER HYDROLASE.
FT CONFLICT 51 52 GS -> AG (IN REF. 2).
FT CONFLICT 90 92 HAR -> RAL (IN REF. 2).
FT CONFLICT 123 123 W -> R (IN REF. 2).
SQ SEQUENCE 453 AA; 49701 MW; F48C2C61475072B2 CRC64;

Query Match 50.9%; Score 59; DB 1; Length 453;
Best Local Similarity 66.7%; Pred. No. 2.31e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 124 AMEPERPLW 132
| | | | |
QY 3 AMDPTPLW 11

RESULT 8
ID NCAP_MACHU STANDARD; PRT; 564 AA.
AC P26578;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NUCLEOCAPSID PROTEIN (NUCLEOPROTEIN).
GN N.
OS Machupo virus.
OC Viruses; SSRNA negative-strand viruses; Arenaviridae; Arenavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AA288-77;
RX MEDLINE; 92296904.
RA Griffiths C., Wilson S.M., Clegg J.C.S.;
RT "Sequence of the nucleocapsid protein gene of Machupo virus: close
RT relationship with another South American pathogenic arenavirus,
RT Junin.";
RL Arch. Virol. 124:371-377(1992).
CC -1- SIMILARITY: BELONGS TO THE ARENAVIRUSES NUCLEOCAPSID PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62616; CAA44486.1; -
DR EMBL; S18042; VHXPMV.
DR PFAM; PF00843; Arena_nucleocap; 1.
KW Nucleocapsid.
SQ SEQUENCE 564 AA; 63300 MW; D34C575B1A188BD8 CRC64;

Query Match 50.9%; Score 59; DB 1; Length 564;
Best Local Similarity 46.7%; Pred. No. 2.31e+00;

Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 368 VKLDPTNLMWDIE 382
:|||||:
Qy 1 LKAMDPTPLWIKTE 15

RESULT 9

ID LAF4_HUMAN STANDARD; PRT; 1227 AA.
AC P51826;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE LAF-4 PROTEIN (LYMPHOID NUCLEAR PROTEIN).
GN LAF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96141096.
RA Ma C., Staudt L.M.;
RT "LAF-4 encodes a lymphoid nuclear protein with transactivation
RT potential that is homologous to AF-4, the gene fused to MLL in
RT t(4;11) leukemias".
RL Blood 87:734-745(1996).
CC -!- FUNCTION: PUTATIVE TRANSCRIPTION ACTIVATOR THAT MAY FUNCTION IN
CC LYMPHOID DEVELOPMENT AND ONCOGENESIS. BINDS, IN VITRO, TO
CC DOUBLE-STRANDED DNA.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN LYMPHOID TISSUES,
CC -!- HIGHEST LEVELS BEING FOUND IN THE THYMUS.
CC -!- SIMILARITY: TO AF4 AND OX19.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U34360; AAA98763.1; -;
KW Nuclear protein; Transcription regulation; Activator; DNA-binding.
FT DOMAIN 413 419 POLY-SER.
FT DOMAIN 422 432 POLY-SER.
FT DOMAIN 440 445 POLY-SER.
FT DOMAIN 670 679 POLY-SER.
SQ SEQUENCE 1227 AA; 133734 MW; 6348896FD7E9BBE7 CRC64;

Query Match 50.9%; Score 59; DB 1; Length 1227;
Best Local Similarity 40.0%; Pred. No. 2.31e+00;

Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 757 LKSDSEIRSLWKID 771
:|||||:
Qy 1 LKAMDPTPLWIKTE 15

RESULT 10

ID YAVJ_RHISN STANDARD; PRT; 351 AA.
AC Q53218;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 39.2 KD PROTEIN YAVJ.
GN YAVJ.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE; 97305956.
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
RN [2]
RP SEQUENCE OF 1-279 FROM N.A.
RX MEDLINE; 96389014.
RA Freiberg C., Perret X., Broughton W.J., Rosenthal A.;
RT "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.
RT NGR234 using dye terminators and a thermostable 'sequenase': a
RT beginning.";
RL Genome Res. 6:590-600(1996).
CC -!- SIMILARITY: TO ALKALINE MONOOXYGENASE ALPHA AND BETA CHAINS (EC
CC 1.14.14.3) (BACTERIAL LUCIFERASE). BUT DISTANTLY RELATED.
CC -!- SIMILARITY: TO Y4WF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; Z68203; CAA92425.1; -;
DR EMBL; AE00101; AAB91898.1; -;
DR PFAM; PF00296; bac_luciferase; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Plasmid.
SQ SEQUENCE 351 AA; 39158 MW; 866BDE3B8A40C88D CRC64;

Query Match 50.0%; Score 58; DB 1; Length 351;
Best Local Similarity 77.8%; Pred. No. 3.53e+00;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 167 PTPPIWIAT 175
:|||||:
Qy 6 PTPPLWIAT 14

RESULT 11

ID YNG2_CAEEL STANDARD; PRT; 492 AA.
AC Q21653;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHEICAL 54.3 KD PROTEIN R02F2.2 IN CHROMOSOME III.
GN R02F2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Pauley A.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HUMAN DS-1 AND YEAST YOL114C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; U00055; AAA50719.1; -;
DR WORMPEP; R02F2.2; CE00687.
KW Hypothetical protein.
SQ SEQUENCE 492 AA; 54302 MW; FBE7514112A8F413 CRC64;

Query Match 50.0%; Score 58; DB 1; Length 492;
Best Local Similarity 54.5%; Pred. No. 3.53e+00;

```
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 101 KRVDSTPSLW 111
   1 :|||:|
QY 2 KAMDPTPLWI 12

RESULT 12
ID RRPLAHSV9 STANDARD; PRT; 1305 AA.
AC 070695;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.48) (VPI).
GN SI.
OS African horse sickness virus 9 (AHSV-9) (African horse sickness virus
   (serotype 9)).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX Vreede F.T., Huismans H.;
RT "Sequence analysis of the RNA polymerase gene of African horse
   sickness virus";
RL Arch. Virol. 143:413-419(1998).
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE
   + RNA(N).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U94887; AAC40586.1;
KW Transferase; Nucleotidyltransferase; RNA-directed RNA polymerase.
SQ SEQUENCE 1305 AA; 150293 MW; 321E9E7F3CF6F11A CRC64;

Query Match 50.0%; Score 58; DB 1; Length 1305;
Best Local Similarity 50.0%; Pred. No. 3.53e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 960 VKLIDQSPPLWNET 973
   1 :|||:|
QY 1 KAMDPTPLWI 14

RESULT 13
ID C04_HUMAN STANDARD; PRT; 1741 AA.
AC P01028;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE COMPLEMENT C4 PRECURSOR [CONTAINS: C4A ANAPHYLATOXIN].
GN C4A AND C4B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE OF 1-22 AND 1056-1225 FROM N.A.
RX MEDLINE; 85156269.
RA Belt K.T., Yu C.Y., Carroll M.C., Porter R.R.;
RT "Polymorphism of human complement component C4.";
RL Immunogenetics 21:173-180(1985).
RN [2]
RP SEQUENCE OF 20-1741 FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 84156344.
RA Belt K.T., Carroll M.C., Porter R.R.;
RT "The structural basis of the multiple forms of human complement
   component C4.";
```

```
RL Cell 36:907-914(1984).
RN [3]
RP SEQUENCE OF 680-756.
RX MEDLINE; 81264286.
RA Moon K.E., Gorski J.P., Hugli T.E.;
RT "Complete primary structure of human C4a anaphylatoxin.";
RL J. Biol. Chem. 256:8685-8692(1981).
RN [4]
RP SEQUENCE OF 957-1044.
RX MEDLINE; 82182029.
RA Campbell R.D., Gagnon J., Porter R.R.;
RT "Amino acid sequence around the thiol and reactive acyl groups of
   human complement component C4.";
RL Biochem. J. 199:359-370(1981).
RN [5]
RP SEQUENCE OF 990-1037.
RX MEDLINE; 82150875.
RA Harrison R.A., Thomas M.L., Tack B.F.;
RT "Sequence determination of the thiolester site of the fourth
   component of human complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:7388-7392(1981).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE; 94282044.
RA Sargent C.A., Anderson M.J., Hsieh S.L., Kendall E.,
   Gomez-Escobar N., Campbell R.D.;
RT "Characterisation of the novel gene G11 lying adjacent to the
   complement C4A gene in the human major histocompatibility complex.";
RL Hum. Mol. Genet. 3:481-488(1994).
RN [7]
RP STRUCTURAL BASIS OF POLYMORPHISM.
RX MEDLINE; 87080272.
RA Yu C.Y., Belt K.T., Giles C.M., Campbell R.D., Porter R.R.;
RT "Structural basis of the polymorphism of human complement components
   C4A and C4B: gene size, reactivity and antigenicity.";
RL EMBO J. 5:2873-2881(1986).
RN [8]
RP VARIANT C4A6 ALLOTYPE.
RX MEDLINE; 92242905.
RA Anderson M.J., Milner C.M., Cotton G.H., Campbell R.D.;
RT "The coding sequence of the hemolytically inactive C4A6 allotype of
   human complement component C4 reveals that a single arginine to
   tryptophan substitution at beta-chain residue 458 is the likely cause
   of the defect.";
RL J. Immunol. 148:2795-2802(1992).
CC -1- FUNCTION: C4 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
   CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM. IT IS PROCESSED BY
   ACTIVATED C1 WHICH REMOVE FROM THE ALPHA CHAIN THE C4A
   ANAPHYLATOXIN.
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C4,
   C4A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
   INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
   PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
   BASOPHILIC LEUKOCYTES.
CC -1- SUBUNIT: THIS PROTEIN IS SYNTHESIZED AS A SINGLE-CHAIN PRECURSOR
   AND, PRIOR TO SECRETION, IS ENZYMATICALLY CLEAVED TO FORM A TRIMER
   OF NONIDENTICAL CHAINS (ALPHA, BETA, AND GAMMA).
CC -1- POLYMORPHISM: HUMAN COMPLEMENT COMPONENT C4 IS POLYMORPHIC WITH AT
   LEAST TWO LOCI, C4A AND C4B. 13 ALLELES OF C4A AND 22 ALLELES OF
   C4B HAVE BEEN DETECTED. THE ALLELE SHOWN HERE IS C4A4.
CC -1- POLYMORPHISM: THE C4A ALLELES CARRY THE BLOOD GROUP RODGERS WHILE
   THE C4B ALLELES CARRY THE BLOOD GROUP CHIDO.
CC -1- DISEASE: THE C4A6 ALLOTYPE IS TOTALLY DEFICIENT IN HEMOLYTIC
   ACTIVITY.
CC -1- MISCELLANEOUS: C4A ALLOYPES REACT MORE RAPIDLY WITH THE AMINO
   GROUP OF PEPTIDE ANTIGENS WHILE C4B ALLOYPES REACT MORE RAPIDLY
   WITH THE HYDROXYL GROUP OF CARBOHYDRATE ANTIGENS.
CC -1- MISCELLANEOUS: C4 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
   PROTEIN.
CC -1- SIMILARITY: TO C3, C5 AND ALPHA-2-MACROGLOBULIN.
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

-----
EMBL; M14823; AAB35617.1; -
DR EMBL; K02403; AAB59537.1; ALT_SEQ.
DR EMBL; M14824; AAB52292.1; -
DR EMBL; X77491; CAA54627.1; -
DR PIR; A01262; C4HU
DR PIR; A17265; A17265.
DR PIR; A29177; A29177.
DR PIR; B20807; B20807.
DR HSSP; P01031; 1KJ5.
DR SWISS-2DPAGE; P01028; HUMAN.
DR MIM; 120790; -
DR MIM; 120810; -
DR MIM; 120820; -
DR PRINTS; PR00004; ANAPHYLATOXIN.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
DR PIR; PF0207; A2M; 1.
KW Complement pathway; Plasma; Glycoprotein; MHC III; Signal;
KW Inflammatory response; Polymorphism; Disease mutation;
KW Blood group antigen.
FT SIGNAL 1 19
FT CHAIN 20 675
FT PROPEP 676 679
FT CHAIN 680 1443
FT PROPEP 1444 1450
FT CHAIN 1451 1741
FT PEPTIDE 680 756
FT DOMAIN 702 736
FT DISULFID 702 728
FT DISULFID 703 735
FT DISULFID 716 736
FT THIOLEST 1010 1013
FT CARBOHYD 226 226
FT CARBOHYD 862 862
FT CARBOHYD 1328 1328
FT CARBOHYD 1391 1391
FT VARIANT 477 477
FT VARIANT 726 726
FT VARIANT 1073 1073
FT VARIANT 1120 1125
FT VARIANT 1176 1176
FT VARIANT 1201 1201
FT VARIANT 1207 1207
FT VARIANT 1210 1210
FT VARIANT 1286 1286
FT CONFLICT 727 727
FT CONFLICT 1013 1013
FT SEQUENCE 1741 AA; 192335 MW; E0164408D6712001 CRC64;
Query Match 50.0%; Score 58; DB 1; Length 1741;
Best Local Similarity 50.0%; Pred. No. 3.53e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
-----

```

Db 1253 PMPQAPALWIET 1264

QY 3 AMDPTPLWIK 14

```

RESULT 14
ID YSCK_YERPS STANDARD; PRT; 209 AA.
AC Q00927;
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE YOP PROTEINS TRANSLOCATION PROTEIN K (LOW CALCIUM RESPONSE LOCUS
DE PROTEIN KB).
GN YSCK OR LCRKB.
OS Yersinia pseudotuberculosis.
OC Plasmid pIB1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YPIII;
RX MEDLINE; 92250432.
RA Rimpilainen M., Forsberg A., Wolf-Watz H.;
RT "A novel protein, LcrO, involved in the low-calcium response of
RT Yersinia pseudotuberculosis shows extensive homology to YopH.";
RL J. Bacteriol. 174:3355-3363(1992).
CC -!- FUNCTION: BELONGS TO AN OPERON INVOLVED IN THE TRANSLLOCATION OF
CC YOP PROTEINS ACROSS THE BACTERIAL MEMBRANES OR IN THE SPECIFIC
CC CONTROL OF THIS FUNCTION.
CC -!- INDUCTION: AT 37 DEGREES CELSIUS IN THE ABSENCE OF CALCIUM.
CC -!- SIMILARITY: HIGH, WITH Y. ENTEROCOLITICA CORRESPONDING ORF.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

-----
DR EMBL; M83986; AAA27652.1; -
KW Plasmid; Virulence.
SQ SEQUENCE 209 AA; 23991 MW; 1E19276045508988 CRC64;
Query Match 49.1%; Score 57; DB 1; Length 209;
Best Local Similarity 53.8%; Pred. No. 5.36e+00;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
-----

```

```

Db 152 LAAMEPOPOQAWCK 164
QY 1 LKAMDPTPLWIK 13
RESULT 15
ID YSCK_YEREN STANDARD; PRT; 209 AA.
AC Q01252;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE YOP PROTEINS TRANSLOCATION PROTEIN K.
GN YSCK.
OS Yersinia enterocolitica.
OC Plasmid pYV.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=439-80 / SEROTYPE O:9;
RX MEDLINE; 91317716.
RA Michiels T., Vancotteghem J.-C., de Rouvoit C., China B., Gustin A.,
RA Boudry P., Cornelis G.R.;
RT "Analysis of virC, an operon involved in the secretion of Yop
RT proteins by Yersinia enterocolitica.";
RL J. Bacteriol. 173:4994-5009(1991).
CC -!- FUNCTION: BELONGS TO AN OPERON INVOLVED IN THE TRANSLLOCATION OF
CC YOP PROTEINS ACROSS THE BACTERIAL MEMBRANES OR IN THE SPECIFIC

```

CC CONTROL OF THIS FUNCTION.
 CC -!- INDUCTION: AT 37 DEGREES CELSIUS IN THE ABSENCE OF CALCIUM.
 CC -!- SIMILARITY: HIGH, WITH Y.PSEUDOTUBERCULOSIS CORRESPONDING ORF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M74011; AAC37028.1; -
 CC PIR: B40049; B40049.
 CC Plasmid; Virulence.
 KW SEQUENCE 209 AA; 23998 MW; 05FEDC449AABA988 CRC64;

Query Match 49.1%; Score 57; DB 1; Length 209;
 Best Local Similarity 53.8%; Pred.No. 5.36e+00;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 Db 152 LAAMEPOPOQAWCK 164
 QY 1 LKAMDPTPLWIK 13

Search completed: Fri Jul 7 15:32:50 2000
 Job time : 9 secs.